

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 05:49:35 ; Search time 1537 Seconds

(Without alignments)  
8994.036 Million cell updates/sec

Title: US-09-719-002-1

Perfect score: 475  
Sequence: 1 gaattcttatgtgcagctcga.....acaacacacacacatcatg 475

Scoring table: OLIGO\_MUC  
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	400	84.2	473	6 AX008564	AX008564 Sequence
2	24	5.1	42	6 AX008582	AX008582 Sequence
3	22	4.6	29	6 AX008576	AX008576 Sequence
4	22	4.6	33	6 AX008574	AX008574 Sequence
5	22	4.6	40	6 AX008579	AX008579 Sequence
6	21	4.6	65361	2 AC091863	AC091863 Homo sapi
7	21	4.4	73218	2 AC121724	AC121724 Rattus no
8	21	4.4	92250	2 AC026693	AC026693 Homo sapi
9	21	4.4	122697	2 CNS08C8E	AL731874 Oryza sat
10	21	4.4	130732	2 AP003848	AP003848 Oryza sat
11	21	4.4	139613	2 CNS08C7T	AL731744 Oryza sat
12	21	4.4	149988	2 AC011929	AC011929 Homo sapi
13	21	4.4	156248	9 AC027322	AC027322 Homo sapi
14	21	4.4	162227	2 AC095224	AC095224 Rattus no
15	20	4.2	29	6 AX008577	AX008577 Sequence
16	20	4.2	29	6 AX008578	AX008578 Sequence
17	20	4.2	29	6 AX008575	AX008575 Sequence
18	20	4.2	918	4 AF271901	AF271901 Bos tauru
19	20	4.2	7595	4 BTCA5K35	X14908 Bovine gene
20	20	4.2	39416	3 CEB0001	Z69634 Caenorhabdl
21	20	4.2	78448	2 AC114836	AC114836 Rattus no
22	20	4.2	99135	2 AC094561	AC094561 Rattus no
23	20	4.2	111656	2 AC010449	AC010449 Homo sapi
24	20	4.2	128676	2 AC126877	AC126877 Rattus no
25	20	4.2	134955	10 AL714026	AL714026 Mouse DNA
26	20	4.2	135619	9 AL554809	AL554809 Human DNA
27	20	4.2	138996	2 AC008803	AC008803 Homo sapi
28	20	4.2	144841	2 AC121995	AC121995 Mus muscu
29	20	4.2	145658	9 AC095154	AC095154 Homo sapi
30	20	4.2	149751	2 AC048339	AC048339 Homo sapi
31	20	4.2	151111	2 AC103077	AC103077 Rattus no
32	20	4.2	157348	2 AP002446	AP002446 Homo sapi
33	20	4.2	162731	2 AC109162	AC109162 Mus muscu
34	20	4.2	164429	9 AC026366	AC026366 Homo sapi
35	20	4.2	166841	2 AC114523	AC114523 Rattus no
36	20	4.2	168960	2 AC015958	AC015958 Homo sapi
37	20	4.2	172189	9 AC109518	AC109518 Homo sapi
38	20	4.2	176224	2 AP000898	AP000898 Homo sapi
39	20	4.2	176999	2 AC053518	AC053518 Homo sapi
40	20	4.2	177444	2 AC084238	AC084238 Mus muscu
41	20	4.2	184436	9 AP000919	AP000919 Homo sapi
42	20	4.2	185178	2 AP000937	AP000937 Homo sapi
43	20	4.2	187310	2 AC101870	AC101870 Mus muscu
44	20	4.2	192519	10 MM0278435	AJ278435 Mus muscu
45	20	4.2	195907	2 AC095185	AC095185 Rattus no

# ALIGNMENTS

RESULT 1  
LOCUS AX008564  
DEFINITION Sequence 1 from Patent WO9966057.  
ACCESSION AX008564  
VERSION AX008564.1 GI:9996114  
KEYWORDS  
ORGANISM  
SOURCE  
garden asparagus.  
Asparagus officinalis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;  
Asparagaceae; Asparagus.  
REFERENCE  
1 (bases 1 to 473)  
Draper, J., Keaton, P., and Paul, W.  
Inducible promoters

JOURNAL Patent: WO 9966057-A 1 23-DEC-1999;  
DRAPER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
WYATT (GB)

FEATURES  
source Location/Qualifiers  
1..473  
/organism="Asparagus officinalis"  
/db\_xref="taxon:4686"

BASE COUNT 162 a 112 c 67 g 132 t

ORIGIN

Query Match 84.2%; Score 400; DB 6; Length 473;  
Best Local Similarity 100.0%; Pred. No. 3.5e-209;  
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATCTTATTCGACCTGACTCTCTGTGTGTCGCGAGGTGCTGCAATTCGT 60  
|||||  
Db 1 GAATCTTATTCGACCTGACTCTCTGTGTGTCGCGAGGTGCTGCAATTCGT 60

QY 61 TCGCGACACACATAGTGGTCTTGTGATTTGACAGTTCCATATTTTCCATGTCAT 120  
|||||  
Db 61 TCGCGACACACATAGTGGTCTTGTGATTTGACAGTTCCATATTTTCCATGTCAT 120

QY 121 GAGAGAGACATAGTCTTAAGTAATTAATCCCTTAATACATACAAAGCAT 180  
|||||  
Db 121 GAGAGAGACATAGTCTTAAGTAATTAATCCCTTAATACATACAAAGCAT 180

QY 181 GACACATCCACAGAAAAATTTCTAATAGTCTTGTGTAAGTAATGGAACATGATAC 240  
|||||  
Db 181 GACACATCCACAGAAAAATTTCTAATAGTCTTGTGTAAGTAATGGAACATGATAC 240

QY 241 CTACATTAATTAACAATTTTGAATTAATTAAGAAAGTTCTAATGTAAGACTAGT 300  
|||||  
Db 241 CTACATTAATTAACAATTTTGAATTAATTAAGAAAGTTCTAATGTAAGACTAGT 300

QY 301 TCTAATGTAAGACTAGTCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 360  
|||||  
Db 301 TCTAATGTAAGACTAGTCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 360

QY 361 CAATCGAATTAATCCCTGACATGACATGACACCATCCAA 400  
|||||  
Db 361 CAATCGAATTAATCCCTGACATGACATGACACCATCCAA 400

RESULT 2  
AX008582 42 bp DNA linear PAT 06-SEP-2000  
LOCUS  
DEFINITION Sequence 19 from Patent WO9966057.  
ACCESSION AX008582  
VERSION AX008582.1 GI:9996132  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial construct.  
REFERENCE 1 (bases 1 to 42)  
AUTHORS Draper,J., Kenton,P. and Paul,W.  
TITLE Inducible promoters  
JOURNAL Patent: WO 9966057-A 19 23-DEC-1999;  
DRAPER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
WYATT (GB)

FEATURES  
source Location/Qualifiers  
1..42  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"

BASE COUNT 13 a 7 c 14 g 8 t

ORIGIN

Query Match 5.1%; Score 24; DB 6; Length 42;  
Best Local Similarity 100.0%; Pred. No. 0.092;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 GTCCAGCAACTGCTACTATTTC 340  
|||||

Db 42 GTCCAGCAACTGCTACTATTTC 19

RESULT 3  
AX008576 29 bp DNA linear PAT 06-SEP-2000  
LOCUS  
DEFINITION Sequence 13 from Patent WO9966057.  
ACCESSION AX008576  
VERSION AX008576.1 GI:9996126  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial construct.  
REFERENCE 1 (bases 1 to 29)  
AUTHORS Draper,J., Kenton,P. and Paul,W.  
TITLE Inducible promoters  
JOURNAL Patent: WO 9966057-A 13 23-DEC-1999;  
DRAPER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
WYATT (GB)

FEATURES  
source Location/Qualifiers  
1..29  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"

BASE COUNT 9 a 7 c 8 g 5 t

ORIGIN

Query Match 4.6%; Score 22; DB 6; Length 29;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 TTCATGTCATGAGAGAGCAGC 131  
|||||  
Db 8 TTCATGTCATGAGAGAGCAGC 29

RESULT 4  
AX008574 33 bp DNA linear PAT 06-SEP-2000  
LOCUS  
DEFINITION Sequence 11 from Patent WO9966057.  
ACCESSION AX008574  
VERSION AX008574.1 GI:9996124  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM artificial construct.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS Draper,J., Kenton,P. and Paul,W.  
TITLE Inducible promoters  
JOURNAL Patent: WO 9966057-A 11 23-DEC-1999;  
DRAPER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
WYATT (GB)

FEATURES  
source Location/Qualifiers  
1..33  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"

BASE COUNT 6 a 10 c 7 g 10 t

ORIGIN

Query Match 4.6%; Score 22; DB 6; Length 33;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TTCTATTGCGACCTGACTCTC 25  
|||||  
Db 12 TTCTATTGCGACCTGACTCTC 33

RESULT 5  
AX008579 40 bp DNA linear PAT 06-SEP-2000  
LOCUS  
DEFINITION Sequence 16 from Patent WO9966057.

ACCESSION AX008579 GI:9996129  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

synthetic construct.  
synthetic construct  
artificial sequences.  
1 (bases 1 to 40)  
Draper, J., Kenton, P. and Paul, W.  
Inducible promoters  
Patent: WO 966057-A 16 23-DEC-1999;  
DRAVER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
WYATT (GB)

FEATURES  
Source  
Location/Qualifiers  
1..40  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="primer"

BASE COUNT 4 a 6 c 14 g 16 t  
ORIGIN

Query Match 4.6%; Score 22; DB 6; Length 40;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Gaps 0;

QY 451 CAAAACACACACACATC 472  
|||||  
Db 40 CAAAACACACACACATC 19

RESULT 6  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC091863 65361 bp DNA linear HTG 09-JUN-2001  
Homo sapiens chromosome 5 clone CTD-235015, WORKING DRAFT SEQUENCE,  
14 unordered pieces.  
AC091863  
AC091863.1 GI:14333799  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 65361)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 65361)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Project Information  
Center Project Name: 762146  
Center Clone Name: CITB-HL\_235015

Summary Statistics  
Consensus quality: 54296 bases at least Q40  
Consensus quality: 56671 bases at least Q30  
Consensus quality: 57432 bases at least Q20  
Estimated insert size: 168010; agarose-1p estimation  
Estimated insert size: 64061; sum-of-ctnigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-1p estimation  
Quality coverage: 14.12 in Q20 bases; sum-of-ctnigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 ctnigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the ctnigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

as soon as it is available and the accession number will  
\* be preserved.  
1 1343: contig of 1343 bp in length  
\* 1344: gap of unknown length  
\* 1444: contig of 1097 bp in length  
\* 2541: gap of unknown length  
\* 2641: gap of 1071 bp in length  
\* 3712: contig of 1010 bp in length  
\* 3812: contig of 1010 bp in length  
\* 4821: gap of unknown length  
\* 4822: contig of 1206 bp in length  
\* 4922: gap of unknown length  
\* 6128: gap of unknown length  
\* 6228: gap of 5683 bp in length  
\* 6228: gap of unknown length  
\* 11911: gap of unknown length  
\* 12011: gap of 2837 bp in length  
\* 14847: gap of unknown length  
\* 14848: gap of 3529 bp in length  
\* 14948: gap of unknown length  
\* 18477: gap of 6517 bp in length  
\* 18577: gap of unknown length  
\* 25093: contig of 6517 bp in length  
\* 25094: gap of unknown length  
\* 25194: gap of 7917 bp in length  
\* 33111: gap of unknown length  
\* 33210: gap of unknown length  
\* 33211: gap of 6585 bp in length  
\* 39795: contig of 6585 bp in length  
\* 39796: gap of unknown length  
\* 39895: gap of unknown length  
\* 39896: gap of 6847 bp in length  
\* 46743: gap of unknown length  
\* 46843: gap of 10980 bp in length  
\* 57823: gap of unknown length  
\* 57922: gap of 7439 bp in length.  
57923

FEATURES  
Source  
Location/Qualifiers  
1..65361  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-235015"  
/clone\_11p="Caltech human BAC library D"

BASE COUNT 19445 a 13085 c 11801 g 19653 t 1377 others  
ORIGIN

Query Match 4.4%; Score 21; DB 2; Length 65361;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Gaps 0;

QY 241 CTACATTATATACACTTTG 261  
|||||  
Db 16620 CTACATTATATACACTTTG 16640

RESULT 7  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC121724 73218 bp DNA linear HTG 24-JUL-2002  
Rattus norvegicus clone CH230-332D19, \*\* SEQUENCING IN PROGRESS  
\*\*\*, 39 unordered pieces.  
AC121724  
AC121724.2 GI:21909372  
HTG; HTGS\_PHASE1.  
Rattus norvegicus.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 73218)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Muzny, D.M., Adams, C., Adio-Oduola, B., Alt-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaralunga, H.C., Aye, J.R., Ayale, M., Banks, T.,  
Barbata, J., Benton, J., Bimge, K., Blankenburg, K., Bonini, D.,  
Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dehorne, S.R., David, R.,  
Davila, M.L., Davis, C., Day-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinu, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

	Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,Y., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Howel,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y.S., Johnson,R., Jolyvet,S., Jouhad,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kravovic,J., Kurshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,I.J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.U., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maneshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massery,E., Mahoney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miney,T., Mitchell,T., Monabhat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S., Oguh,M., Okunolu,G., Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickett,R., Plante,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojals,A., Rojubenok,I., Rolfe,M., Ruiz,S., Seavey,G., Scheer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonlake,T., Sparks,A., Stanley,H., Stone,H., Tang,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Taniguchi,T., Taylor,C., Taylor,T., Teiford,B., Thomas,N., Thomas,S., Usmali,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallingron,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D., Weinstein,G. and Gibbs,R.	JOURNAL	TITLE	REFERENCE	AUTHORS	-TITLE	JOURNAL
	Worley,K.C. Submitted (24-Jul-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA On Jul 19, 2002 this sequence version replaced gi:21039660.	Genome Center	Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu Project Information Center project name: GXJC Center clone name: CH230-332D19 Sequencing vector: Plasmid: Chemistry: Dye terminator Big Dye: 100% of reads Assembly program: PHRAP; Version 0.990329 Consensus quality: 30431 bases at least Q40 Consensus quality: 35017 bases at least Q20	* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html). NOTE: This is a working draft sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	I 1598 1798 2976	contig of 1698 bp in length gap of unknown length contig of 1178 bp in length	
*	2977	3076:	gap of unknown length	*	1598	contig of 1698 bp in length	
*	3077	4713:	contig of 1637 bp in length	*	1798	gap of unknown length	
*	4714	4813:	gap of unknown length	*	2976	contig of 1178 bp in length	
*	4814	6354:	contig of 1541 bp in length	*			
*	6355	6454:	gap of unknown length	*			
*	6455	7735:	contig of 1281 bp in length	*			
*	7736	7835:	gap of unknown length	*			
*	7836	8905:	contig of 1070 bp in length	*			
*	8906	9005:	gap of unknown length	*			
*	9006	10662:	contig of 1656 bp in length	*			
*	10662	10761:	gap of unknown length	*			
*	10762	12131:	gap of 1370 bp in length	*			
*	12132	12231:	gap of unknown length	*			
*	12232	13749:	contig of 1518 bp in length	*			
*	13750	13849:	gap of unknown length	*			
*	13850	15819:	contig of 1970 bp in length	*			
*	15820	15919:	gap of unknown length	*			
*	15920	16935:	contig of 1016 bp in length	*			
*	16936	17035:	gap of unknown length	*			
*	17036	18637:	contig of 1602 bp in length	*			
*	18638	18737:	gap of unknown length	*			
*	18738	19971:	contig of 1234 bp in length	*			
*	19972	20071:	gap of unknown length	*			
*	20072	21378:	contig of 1307 bp in length	*			
*	21379	21478:	gap of unknown length	*			
*	21479	22528:	contig of 1050 bp in length	*			
*	22529	22628:	gap of unknown length	*			
*	22630	24018:	contig of 1350 bp in length	*			
*	24019	24118:	gap of unknown length	*			
*	24119	25622:	contig of 1504 bp in length	*			
*	25623	25722:	gap of unknown length	*			
*	25723	26949:	contig of 1227 bp in length	*			
*	26950	27049:	gap of unknown length	*			
*	27050	28440:	contig of 1331 bp in length	*			
*	28441	28540:	gap of unknown length	*			
*	28541	29872:	contig of 1332 bp in length	*			
*	29873	29972:	gap of unknown length	*			
*	29973	31074:	contig of 1102 bp in length	*			
*	31075	31174:	gap of unknown length	*			
*	31175	33423:	contig of 2249 bp in length	*			
*	33424	33523:	gap of unknown length	*			
*	33524	35399:	contig of 1876 bp in length	*			
*	35399	35499:	gap of unknown length	*			
*	35400	37239:	contig of 1740 bp in length	*			
*	37240	37339:	gap of unknown length	*			
*	37340	38482:	contig of 1143 bp in length	*			
*	38483	38582:	gap of unknown length	*			
*	38583	40363:	contig of 1781 bp in length	*			
*	40364	40463:	gap of unknown length	*			
*	40464	41882:	contig of 1419 bp in length	*			
*	41883	41982:	gap of unknown length	*			
*	41983	43509:	contig of 1527 bp in length	*			
*	43510	43609:	gap of unknown length	*			
*	43610	45690:	contig of 2081 bp in length	*			
*	45691	45790:	gap of unknown length	*			
*	45791	47013:					



FEATURES \* 68950 73218: contig of 4269 bp in length.  
Location/Qualifiers  
1. 73218  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-332D19"

BASE COUNT 20948 a 15112 c 14120 g 19150 t 3888 others  
ORIGIN

Query Match 4.4%; Score 21; DB 2; Length 73218;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 250 TTACACTTTGCAATATAA 270  
|||||  
DB 27829 TTACACTTTGCAATATAA 27849

RESULT 8  
AC026693 92250 bp DNA linear PRI 03-OCT-2001  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTC-261E10, complete sequence.  
VERSION AC026693.5 GI:15887318  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 92250)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 92250)  
TITLE DOE Joint Genome Institute.  
AUTHORS Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 92250)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
JOURNAL Submitted (09-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 92250)  
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
REFERENCE Submitted (03-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
AUTHORS On Oct 3, 2001 this sequence version replaced g1:10044353.  
TITLE Draft Sequence Produced by DOE Joint Genome Institute  
JOURNAL www.jgi.doe.gov  
www.shgc.stanford.edu  
Qualify: Phrap Quality >=40 99.5% of Sequence:  
Estimated Total Number of Errors is 0.4.  
Location/Qualifiers  
1. 92250  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-261E10"

BASE COUNT 28028 a 17238 c 17152 g 29832 t  
ORIGIN

Query Match 4.4%; Score 21; DB 9; Length 92250;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 CTACATTAATTACAACTTTG 261  
|||||  
DB 58285 CTACATTAATTACAACTTTG 58305

RESULT 9  
CNS08C8E/c 122697 bp DNA linear HTG 29-APR-2002  
LOCUS  
DEFINITION Oryza sativa chromosome 12 clone OSJNB0077C18, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 4 ordered pieces.  
ACCESSION AL731874.1 GI:20372824  
VERSION AL731874  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_ACTIVEFIN.  
SOURCE Oryza sativa.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
REFERENCE 1 (bases 1 to 122697)  
AUTHORS Choisne,N., Orjeda,G., Catolico,L., Demange,N., Wincker,P.,  
Seguene,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,  
Weissenbach,J. and Queller,F.  
Oryza sativa chromosome 12 sequencing  
JOURNAL Unpublished  
TITLE 2 (bases 1 to 122697)  
AUTHORS Genoscope.  
REFERENCE Direct Submission  
JOURNAL Submitted (29-APR-2002) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence.  
Work on the sequence is in progress and the release of this data is  
based on the understanding that the sequence may change as work  
continue. The sequence may be contaminated with foreign sequence  
from E.coli, yeast, vector, phage, etc.  
Contigs composition :  
11793 bp contig from 1 to 11793  
2451 bp contig from 11894 to 14344  
11789 bp contig from 14445 to 26233  
96364 bp contig from 26334 to 122697.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 11793: contig of 11793 bp in length  
\* 11794 11893: gap of 100 bp  
\* 11894 14344: contig of 2451 bp in length  
\* 14345 14444: gap of 100 bp  
\* 14445 26233: contig of 11789 bp in length  
\* 26234 26333: gap of 100 bp  
\* 26334 122697: contig of 96364 bp in length.  
Location/Qualifiers  
1. 122697  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="12"  
/clone="OSJNB0077C18"  
/clone\_1fb="OSJNB"

BASE COUNT 35100 a 27094 c 26028 g 34171 t 304 others  
ORIGIN

Query Match 4.4%; Score 21; DB 2; Length 122697;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 131 CAGGACTAAGTAATTAGCTT 151  
|||||  
DB 100162 CAGGACTAAGTAATTAGCTT 100142

RESULT 10  
AP003848/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
SOURCE  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
OY  
DB

AP003848 130732 bp DNA linear HTG 21-MAR-2002  
Oryza sativa (Japonica cultivar-group) chromosome 7 clone  
OJ1715\_A07, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
AP003848  
AP003848.1 GI:14595193  
HTG: HTGS\_PHASE2.  
Oryza sativa (Japonica cultivar-group) (cultivar:Nipponbare) DNA.  
clone:OJ1715\_A07.  
Oryza sativa (Japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehretaceae; Oryzae; Oryza.  
1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare(G43) genomic DNA, chromosome 7, BAC  
clone:OJ1715\_A07  
Published Only in Database (2001)  
2 (bases 1 to 130732)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (03-JUL-2001) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@agr.affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/,  
Tel:81-298-38-7441, Fax:81-298-38-7468)  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 130732  
/organism="Oryza sativa (Japonica cultivar-group)"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="7"  
/clone="OJ1715\_A07"

36776 a 28066 c 27679 g 38160 t 51 others

131 CATGACTAAAGTAATTAGCTT 151  
|||||  
78724 CATGACTAAAGTAATTAGCTT 78704

RESULT 11  
CNS08C7  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

CNS08C7 139613 bp DNA linear HTG 15-APR-2002  
Oryza sativa chromosome 12 clone OSJNBa0026C14, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.  
AL731744  
AL731744.1 GI:20160279  
HTG: HTGS\_PHASE2; HTGS\_ACTIVEFIN.  
Oryza sativa  
Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehretaceae; Oryzae; Oryza.  
1 (bases 1 to 139613)  
Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
SOURCE  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
OY  
DB

Segrens, B., Pelletier, E., Searpelli, C., Salanoubat, M.,  
Weissenbach, J. and Queller, F.  
Oryza sativa chromosome 12 sequencing  
Unpublished  
2 (bases 1 to 139613)  
Genoscope.  
Direct Submission  
Submitted (15-APR-2002) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence.  
Work on the sequence is in progress and the release of this data is  
based on the understanding that the sequence may change as work  
continue. The sequence may be contaminated with foreign sequence  
from E.coli, yeast, vector, phage, etc.  
Contigs composition :  
3786 bp contig from 1 to 3786  
4831 bp contig from 3887 to 8717  
3718 bp contig from 8818 to 12535  
4235 bp contig from 12636 to 16870  
6706 bp contig from 16921 to 23576  
12914 bp contig from 23777 to 36690  
2950 bp contig from 36791 to 39740  
18336 bp contig from 39841 to 58176  
18030 bp contig from 58277 to 76306  
12219 bp contig from 76407 to 88625  
2501 bp contig from 88726 to 91226  
2193 bp contig from 91327 to 93519  
2004 bp contig from 93630 to 95632  
11756 bp contig from 95724 to 107479  
32034 bp contig from 107580 to 139613.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1. 139613  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="Japonica"  
/db\_xref="taxon:4530"  
/chromosome="12"  
/clone="OSJNBa0026C14"  
/clone\_11b="OSJNBa"

40605 a 23612 c 29656 g 38336 t 1404 others

131 CATGACTAAAGTAATTAGCTT 151  
|||||  
136938 CATGACTAAAGTAATTAGCTT 136958

RESULT 12  
AC011929  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AC011929 149988 bp DNA linear HTG 12-MAR-2000  
Homo sapiens clone RP11-16B11, WORKING DRAFT SEQUENCE, 7 unordered  
pieces.  
AC011929  
AC011929.3 GI:7107941  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 149988)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-16B11  
Unpublished

```

REFERENCE
  2 (bases 1 to 149988)
AUTHORS
  Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
  Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boucknight, B.,
  Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
  Cooke, P., Dearlano, K., Dewar, K., Domino, R., Donnell, L., Doyle, M.,
  Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D., Hooton, L.,
  Hovland, J., Johnson, R., Jones, C., Kahn, L., Karas, A., Klein, J.,
  Leoczky, J., Liu, C., Locke, K., MacDonald, P., Marquis, N.,
  McMan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
  Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
  Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
  Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
  Testa, S., Tirell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X.,
  Wynan, D., Ye, W.J., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
  Submitted (16-Oct-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Feb 28, 2000 this sequence version replaced g1:6453978.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L3509
Center clone name: 16_B_11
-----
Summary Statistics
Sequencing Vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 125296 bases at least Q40
Consensus quality: 146180 bases at least Q30
Insert size: 148000; agarose-fp
Quality coverage: 4.8 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
-----
1 2: contig of 2 bp in length
3 102: gap of 100 bp
103 4317: contig of 4115 bp in length
4318 4317: gap of 100 bp
4318 11679: contig of 7362 bp in length
11680 11779: gap of 100 bp
11780 26507: contig of 14728 bp in length
26508 26607: gap of 100 bp
26608 60280: contig of 33673 bp in length
60281 60380: gap of 100 bp
60381 103134: contig of 42754 bp in length
103135 103234: gap of 100 bp
103235 149988: contig of 46754 bp in length.
-----
Location/Qualifiers
1. 149988
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="RP11-16B11"
/clone_lib="RPC1-11 Human Male BAC"
1. 2
/note="assembly_fragment"
clone_end:17
vector_side:left"
-----
FEATURES
SOURCE
misc_feature
103..4217
/note="assembly_fragment"
4318..11679
/note="assembly_fragment"
11780..26507
/note="assembly_fragment"
26608..60280
/note="assembly_fragment"
60381..103134
/note="assembly_fragment"
103235..149988
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
-----
BASE COUNT
49224 a 29923 c 29730 g 47371 t
ORIGIN
Query Match 4.4% Score 21; DB 2: Length 149988;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
RESULT 13
AC027322/c 156248 bp DNA 1linear PRI 06-DEC-2001
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone CTD-2027610, complete sequence.
ACCESSION
AC027322
VERSION
AC027322.4 GI:17386250
KEYWORDS
HTG.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156248)
DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE
Direct Submission
JOURNAL
Unpublished
2 (bases 1 to 156248)
DOE Joint Genome Institute.
AUTHORS
Direct Submission
JOURNAL
Submitted (06-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Dec 6, 2001 this sequence version replaced g1:9256705.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
-----
Location/Qualifiers
1. 156248
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2027610"
-----
BASE COUNT
49224 a 29923 c 29730 g 47371 t
ORIGIN
Query Match 4.4% Score 21; DB 2: Length 156248;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
241 CTACATTAAATACACTTTG 261
|||||

```

DB 124126 CTACATTAATTACACTTTTG 124106

RESUBMIT 14  
AC095224

DEFINITION AC095224 162227 bp DNA linear HTG 12-JUL-2002  
Rattus norvegicus clone CH230-9M19, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
77 unordered pieces.

ACCESSION AC095224  
VERSION AC095224.3 GI:21728983  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 162227)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,  
Alshrocks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Ditch,H.H.,  
Doutharte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Eatonhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gottell,D.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,J.F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Krtovic,J., Kutesh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtenage,O., Lien,C., Liu,J., Liu,W., Louieged,H.,  
Lohado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S.,  
Moser,N., Neal,D., Newton,J., Newton,S., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunnu,G.,  
Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Slason,I.,  
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansel,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,  
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Wernstock,G. and Gibbs,R.

TITLE Unpublished

JOURNAL Direct Submission  
REFERENCE 2 (bases 1 to 162227)

AUTHORS Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 162227)  
Worley,K.C.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jul 11, 2002 this sequence version replaced g1:1942456.  
Genome Center

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GPKB  
Center clone name: CH230-9M19  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 95945 bases at least Q40  
Consensus quality: 100752 bases at least Q30  
Consensus quality: 104634 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 77 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
1013: contig of 1013 bp in length  
1014  
1113: gap of unknown length  
1114  
2116: contig of 1003 bp in length  
2117  
2216: gap of unknown length  
2217  
3332: contig of 1116 bp in length  
3333  
3432: gap of unknown length  
3433  
4554: contig of 1422 bp in length  
4555  
6233: gap of unknown length  
6234  
6333: gap of unknown length  
6334  
7408: contig of 1075 bp in length  
7409  
7508: gap of unknown length  
7509  
8753: contig of 1245 bp in length  
8754  
8853: gap of unknown length  
8854  
9858: contig of 1005 bp in length  
9859  
9958: gap of unknown length  
11088: contig of 1130 bp in length  
11089  
11188: gap of unknown length  
11189  
12543: contig of 1355 bp in length  
12544  
12543: gap of unknown length  
12544  
14118: contig of 1475 bp in length  
14119  
14218: gap of unknown length  
14219  
15523: contig of 1305 bp in length  
15524  
15623: gap of unknown length  
15624  
16842: contig of 1219 bp in length  
16843  
16842: gap of unknown length  
16843  
18128: contig of 1186 bp in length  
18129  
18328: gap of unknown length  
18329  
19393: contig of 1065 bp in length  
19394  
19393: gap of unknown length  
19394  
20487: contig of 1094 bp in length  
20488  
20587: gap of unknown length  
20588  
21906: contig of 1319 bp in length  
21907  
22006: gap of unknown length  
22007  
23478: contig of 1472 bp in length  
23479  
23578: gap of unknown length  
23579  
25150: contig of 1572 bp in length  
25151  
25250: gap of unknown length  
25251  
27091: contig of 1841 bp in length  
27092  
27191: gap of unknown length  
27192  
29107: contig of 1916 bp in length  
29108  
29207: gap of unknown length  
29209  
30254: contig of 1047 bp in length  
30255  
30354: gap of unknown length  
30355  
31476: contig of 1122 bp in length  
31477  
31576: gap of unknown length  
31577  
32793: contig of 1217 bp in length  
32794  
32893: gap of unknown length

*	32894	34419:	contig of 1526 bp in length
*	34420	34519:	gap of unknown length
*	34520	35690:	contig of 1171 bp in length
*	35691	35790:	gap of unknown length
*	35791	37618:	contig of 1828 bp in length
*	37619	37718:	gap of unknown length
*	37719	39685:	contig of 1967 bp in length
*	39686	39785:	gap of unknown length
*	42058	42057:	contig of 2272 bp in length
*	42158	42157:	gap of unknown length
*	43804	43803:	contig of 1646 bp in length
*	43904	43903:	gap of unknown length
*	45161	45160:	contig of 1257 bp in length
*	45261	45260:	gap of unknown length
*	46530	46529:	contig of 1269 bp in length
*	46630	46629:	gap of unknown length
*	48239	48238:	contig of 1609 bp in length
*	48339	48338:	gap of unknown length
*	49340	49339:	contig of 1601 bp in length
*	50040	50039:	gap of unknown length
*	51457	51456:	contig of 1417 bp in length
*	51557	51556:	gap of unknown length
*	53593	53592:	contig of 2036 bp in length
*	53693	53692:	gap of unknown length
*	54943	54942:	contig of 1250 bp in length
*	55043	55042:	gap of unknown length
*	56369	56368:	contig of 1326 bp in length
*	56469	56468:	gap of unknown length
*	59085	59084:	contig of 2616 bp in length
*	59185	59184:	gap of unknown length
*	60898	60897:	contig of 1713 bp in length
*	60998	60997:	gap of unknown length
*	62163	62162:	contig of 1165 bp in length
*	62263	62262:	gap of unknown length
*	64036	64035:	contig of 1773 bp in length
*	64136	64135:	gap of unknown length
*	66275	66274:	contig of 2139 bp in length
*	66375	66374:	gap of unknown length
*	67901	67900:	contig of 1526 bp in length
*	68001	68000:	gap of unknown length
*	69828	69827:	contig of 1827 bp in length
*	69928	69927:	gap of unknown length
*	71108	71107:	contig of 1180 bp in length
*	72618	72617:	contig of 1410 bp in length
*	72718	72717:	gap of unknown length
*	73819	73818:	contig of 1101 bp in length
*	73919	73918:	gap of unknown length
*	75745	75744:	contig of 1826 bp in length
*	75845	75844:	gap of unknown length
*	77439	77438:	contig of 1594 bp in length
*	77539	77538:	gap of unknown length
*	80160	80159:	contig of 2621 bp in length
*	80260	80259:	gap of unknown length
*	82281	82280:	contig of 2021 bp in length
*	82381	82380:	gap of unknown length
*	84157	84156:	contig of 1776 bp in length
*	84257	84256:	gap of unknown length
*		86850:	contig of 2594 bp in length

Query Match 4.4%; Score 21; DB 2; Length 162227;  
 Best Local Similarity 100.0%; Pred. No. 4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

=QY 250 TTGAACTTTGCAATATAA 270  
 Db 63077 TTGAACTTTGCAATATAA 63097

RESULT 15  
 AX008577  
 LOCUS AX008577 29 bp DNA linear PAT 06-SEP-2000  
 DEFINITION Sequence 14 from Patent WO966057.

ACCESSION AX008577 GI:9996127  
 VERSION AX008577.1  
 KEYWORDS synthetic construct.  
 SOURCE synthetic construct.  
 ORGANISM artificial sequences.  
 REFERENCE 1 (bases 1 to 29)  
 AUTHORS Draper, J., Kenton, P. and Paul, W.  
 TITLE Inducible promoters  
 JOURNAL Patent: WO 966057-A 14 23-DEC-1999;  
 DRAPER JOHN (GB); KENTON PAUL (GB); BIOGENMA UK LTD (GB); PAUL  
 MYATT (GB)  
 FEATURES Location/Qualifiers  
 source 1..29  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="primer"  
 BASE COUNT 10 a 6 c 6 g 7 t  
 ORIGIN  
 Query Match 4.2%; Score 20; DB 6; Length 29;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 226 TTGAACTGAACTACTACA 245  
 Db 10 TTGAACTGAACTACTACA 29

Search completed: July 20, 2003, 06:50:59  
 Job time: 1541 secs



```

FT      /*tag= d
FT      /note= "Homologous to Carrot PR-3 and PR-4 promoter
FT      sequence"
FT      repeat_region 281..318
FT      /tag= e
FT      /rpt_type= TANDEM
FT      repeat_unit 281..238
FT      /tag= f
FT      /note= "18 bp repeat"
FT      /tag= 409..416
FT      DATA_signal
FT      /tag= 9

```

W09966057-A2.

23-DEC-1999.

21-JUN-1999; 99WO-GB01949.

19-JUN-1998; 98GB-0013345.

(BIOG-) BIOGENMA UK LTD.

Draper J, Kenton P, Paul W;

WPI; 2000-106107/09.

Novel promoters used to control the expression of heterologous genes in transformed plants -

Claim 4; Fig 6; 67pp; English.

The present DNA sequence is a novel inducible promoter, derived from *Asparagus officinalis* thaumatin-like PR-5 related gene, AOPRT-L. The promoter sequence has similarities with other PR promoters. The promoter is responsive to low levels of an environmentally-acceptable and non-phytoxic inducing agent, like Salicylic acid or BTH. The promoters also exhibit low levels of pathogen induced systemic activation and environmentally or developmentally induced expression. The inducible promoter sequence is used to control the expression of heterologous genes in transformed plants, especially genes whose products affect a trait of the plant, such as pathogen resistance, disease control, sterility, fertility or fruit ripening.

Sequence 475 BP; 163 A; 113 C; 67 G; 132 T; 0 other;

Query Match 100.0%; Score 475; DB 21; Length 475;

Best Local Similarity 100.0%; Pred. No. 7,7e-236;

Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 GAATTCCTATGCGACCTGACTCTCTGTGTGCTGCGGAGGTGCTGCAATTCTGT 60
DB      1 GAATTCCTATGCGACCTGACTCTCTGTGTGCTGCGGAGGTGCTGCAATTCTGT 60
QY      61 TGGCCACACATAGTGTGCTGCTGTGATTTGACAGTTCCAAATATTTCCATGTCTAT 120
DB      61 TGGCCACACATAGTGTGCTGCTGTGATTTGACAGTTCCAAATATTTCCATGTCTAT 120
QY      121 GAGGAGGACCATGACATTAAGTATAGCTTAATCCCTAAACTCAATCAACGAGAT 180
DB      121 GAGGAGGACCATGACATTAAGTATAGCTTAATCCCTAAACTCAATCAACGAGAT 180
QY      181 GACACATCCACAGAAAAATTCCTAATAGTCTTGTGCTGTAGAAATGGAAGTGAATAC 240
DB      181 GACACATCCACAGAAAAATTCCTAATAGTCTTGTGCTGTAGAAATGGAAGTGAATAC 240
QY      241 CTACATTAATTAACAATTTTGCAATATAATAAAGAGTTCAACATGAAGACTAGT 300
DB      241 CTACATTAATTAACAATTTTGCAATATAATAAAGAGTTCAACATGAAGACTAGT 300
QY      301 TCTACATGAAGACTAGTCCAGCACTGCTACCTATATCCACAAGGCTTAGACTTCCA 360
DB      301 TCTACATGAAGACTAGTCCAGCACTGCTACCTATATCCACAAGGCTTAGACTTCCA 360

```

```

QY      361 CAATGAGATTATCCATGAGCTAGTGACACCATCCAAATATCCATTAATACCTG 420
DB      361 CAATGAGATTATCCATGAGCTAGTGACACCATCCAAATATCCATTAATACCTG 420
QY      421 CCCATCCCCCTCCCTCCAGACATCATCTAATCCAAAACCAACACCAACGATG 475
DB      421 CCCATCCCCCTCCCTCCAGACATCATCTAATCCAAAACCAACACCAACGATG 475

```

RESULT 2

AAZ29523/C

AAZ29523; standard; DNA; 42 BP.

14-MAR-2000 (first entry)

Primer-4 used for construction of AOPRT-Lx3 promoter.

Inducible promoter; Thaumatin-like PR-5 related gene; AOPRT-L; primer;

non-phytoxic inducing agent; Salicylic acid; SA, BTH, environmental;

developmental; GUS construct; multimerisation; SA responsive element;

systemic activation; AOPRT-Lx3 promoter; p22-JIT60; pUC19; ss.

Synthetic.

W09966057-A2.

23-DEC-1999.

21-JUN-1999; 99WO-GB01949.

19-JUN-1998; 98GB-0013345.

(BIOG-) BIOGENMA UK LTD.

Draper J, Kenton P, Paul W;

WPI; 2000-106107/09.

Novel promoters used to control the expression of heterologous genes in transformed plants -

Example 12; Page 41; 67pp; English.

The present DNA sequence is a PCR primer-4, used for the construction of the AOPRT-Lx3 promoter. This primer is used to amplify the region from -133 bp to -247 bp of the AOPRT-L promoter, from p22-JIT60 and cloned into pUC19.

Sequence 42 BP; 13 A; 7 C; 14 G; 8 T; 0 other;

Query Match 5.1%; Score 24; DB 21; Length 42;

Best Local Similarity 100.0%; Pred. No. 0.028;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      317 GTCCAGAGCTGTAACCTATATCC 340
DB      42 GTCCAGAGCTGTAACCTATATCC 19

```

RESULT 3

AAZ29516

AAZ29516; standard; DNA; 29 BP.

14-MAR-2000 (first entry)

Primer-1 for identification of SA responsive element in AOPRT-L promoter.

Inducible promoter; Thaumatin-like PR-5 related gene; AOPRT-L; primer;

non-phytoxic inducing agent; Salicylic acid; SA, BTH, environmental;

KW developmental; GUS construct; multimerisation; SA responsive element;  
XX systemic activation; Inverse PCR; PCR; ss.  
XX Synthetic.  
XX WO966057-A2.  
XX PN 23-DEC-1999.  
XX PD  
XX PF 21-JUN-1999; 99WO-GB01949.  
XX PR 19-JUN-1998; 98GB-0013345.  
XX PA (BIOG-) BIOGEMMA UK LTD.  
XX PI Draper J, Kenton P, Paul W;  
XX WPI, 2000-106107/09.  
XX PT Novel promoters used to control the expression of heterologous genes in  
XX transformed plants  
XX Example 12; Page 40; 67pp; English.  
XX PS  
XX CC The present DNA sequence is a PCR primer-1, used for the identification  
XX CC and multimerisation of a salicylic acid, SA/BTH responsive element in  
XX CC the AOPRT-L promoter region. This primer is designed to regions of the  
XX CC AOPRT-L promoter and used along with PCR primer-4 for the construction  
XX CC of GUS fusion constructs.  
XX SQ Sequence 29 BP; 9 A; 7 C; 8 G; 5 T; 0 other;  
Query Match 4.6%; Score 22; DB 21; Length 29;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 110 TTCATGTCATGAGAGAGAC 131  
Db 8 TTCATGTCATGAGAGAGAC 29  
RESULT 4  
AAZ29514  
ID AAZ29514 standard; DNA; 33 BP.  
XX AC AAZ29514;  
XX DT 14-MAR-2000 (first entry)  
XX DE Primer-1 for construction of AOPRT-L promoter-GUS chimeric gene.  
XX KM Inducible promoter; Thaumatin-like PR-5 related gene; AOPRT-L; primer;  
XX KM non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;  
XX KM systemic activation; developmental; pIPCR-TA; Inverse PCR; IPCR; ss.  
XX OS Synthetic.  
XX PN WO966057-A2.  
XX PD 23-DEC-1999.  
XX PF 21-JUN-1999; 99WO-GB01949.  
XX PR 19-JUN-1998; 98GB-0013345.  
XX PA (BIOG-) BIOGEMMA UK LTD.  
XX PI Draper J, Kenton P, Paul W;  
XX WPI, 2000-106107/09.  
XX PT Novel promoters used to control the expression of heterologous genes in  
XX transformed plants

XX XX Example 3; Page 32; 67pp; English.  
XX PS  
XX CC The present DNA sequence is a PCR primer-1, used for the isolation of  
XX CC AOPRT-L promoter region. The promoter sequence was obtained from pIPCR-TA  
XX CC using this primer. This primer is designed against both the 5' and 3' ends  
XX CC of the promoter, with extensions to provide appropriate restriction  
XX CC sites for further cloning.  
XX SQ Sequence 33 BP; 6 A; 10 C; 7 G; 10 T; 0 other;  
Query Match 4.6%; Score 22; DB 21; Length 33;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 TTCTATTGGGACCTGACTCTC 25  
Db 12 TTCTATTGGGACCTGACTCTC 33  
RESULT 5  
AAZ29519/c  
ID AAZ29519 standard; DNA; 40 BP.  
XX AC AAZ29519;  
XX DT 14-MAR-2000 (first entry)  
XX DE Primer-4 for identification of SA responsive element in AOPRT-L promoter.  
XX KM Inducible promoter; Thaumatin-like PR-5 related gene; AOPRT-L; primer;  
XX KM non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;  
XX KM developmental; GUS construct; multimerisation; SA responsive element;  
XX KM systemic activation; template; p22-JIT60; Inverse PCR; IPCR; ss.  
XX OS Synthetic.  
XX PN WO966057-A2.  
XX PD 23-DEC-1999.  
XX PF 21-JUN-1999; 99WO-GB01949.  
XX PR 19-JUN-1998; 98GB-0013345.  
XX PA (BIOG-) BIOGEMMA UK LTD.  
XX PI Draper J, Kenton P, Paul W;  
XX WPI, 2000-106107/09.  
XX PT Novel promoters used to control the expression of heterologous genes in  
XX PT transformed plants  
XX PS Example 12; Page 40; 67pp; English.  
XX CC The present DNA sequence is a PCR primer-4, used for the identification  
XX CC and multimerisation of a salicylic acid, SA/BTH responsive element in  
XX CC the AOPRT-L promoter region. This primer is used together with PCR  
XX CC primers 1-3, using p22-JIT60 as template, for the construction of  
XX CC GUS fusion constructs.  
XX SQ Sequence 40 BP; 4 A; 6 C; 14 G; 16 T; 0 other;  
Query Match 4.6%; Score 22; DB 21; Length 40;  
Best Local Similarity 100.0%; Pred. No. 0.3;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 451 CAAAACACACACACCAATC 472  
Db 40 CAAAACACACACACCAATC 19



PF 21-JUN-1999; 99WO-GB01949.  
 PR 19-JUN-1998; 98GB-0013345.  
 PA (BIOG-) BIOGENMA UK LTD.  
 PI Draper J, Kenton P, Paul W;  
 DR WPI; 2000-106107/09.  
 PT Novel promoters used to control the expression of heterologous genes in  
 transformed plants -  
 PS Example 12; Page 40; 67pp; English.  
 CC The present DNA sequence is a PCR primer-2, used for the identification  
 CC and multimerisation of a salicylic acid, SA/BTH responsive element in  
 CC the AOPRT-L promoter region. This primer is designed to regions of  
 CC AOPRT-L promoter and used along with PCR primer-4 for the construction  
 CC of GUS fusion constructs.  
 XX  
 XX Sequence 29 BP; 10 A; 6 C; 6 G; 7 T; 0 other;  
 OY 226 TTGGAACTGATACCTACA 245  
 DB 10 TTGGAACTGATACCTACA 29  
 RESULT 8  
 ID AA229518  
 ID AA229518 standard; DNA; 29 BP.  
 AC AA229518;  
 DT 14-MAR-2000 (first entry)  
 DE Primer-3 for identification of SA responsive element in AOPRT-L promoter.  
 XX Inducible promoter; Thaumatin-like PR-5 related gene; AOPRT-L; primer;  
 KW non-phytoxic inducing agent; salicylic acid; SA; BTH; environmental;  
 KW developmental; GUS construct; multimerisation; SA responsive element;  
 KW systemic activation; Inverse PCR; IPCR; ss.  
 XX Synthetic.  
 OS  
 XX  
 XX WO9966057-A2.  
 PN  
 XX 23-DEC-1999.  
 PD  
 XX 21-JUN-1999; 99WO-GB01949.  
 PF  
 XX 19-JUN-1998; 98GB-0013345.  
 PR  
 XX (BIOG-) BIOGENMA UK LTD.  
 PA  
 XX Draper J, Kenton P, Paul W;  
 PI  
 XX WPI; 2000-106107/09.  
 DR  
 XX Novel promoters used to control the expression of heterologous genes in  
 PT transformed plants -  
 PS Example 12; Page 40; 67pp; English.  
 CC The present DNA sequence is a PCR primer-3, used for the identification  
 CC and multimerisation of a salicylic acid, SA/BTH responsive element in  
 CC the AOPRT-L promoter region. This primer is designed to regions of  
 CC AOPRT-L promoter and used along with PCR primer-4 for the construction  
 CC of GUS fusion constructs.

```
XX      Sequence 29 BP; 9 A; 7 C; 6 G; 7 T; 0 other;
SQ
Query Match      4.2%; Score 20; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 3,3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      341 ACAAGGCTTAGACTTCCA 360
      |||||||
Db      10 ACAAGGCTTAGACTTCCA 29

RESULT 9
AA229515/c
ID      AA229515 standard; DNA; 41 BP.
XX
AC      AA229515;
XX
DT      14-MAR-2000 (first entry)
XX
DE      Primer-2 for construction of AopRT-L promoter-GUS chimeric gene.
XX
KM      Inducible promoter; Thaumatin-like PR-5 related gene; AopRT-L; primer;
KM      non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KM      systemic activation; development; pPCR-TA; Inverse PCR; IPCR; ss.
OS      Synthetic.
XX
PN      WO9966057-A2.
XX
PD      23-DEC-1999.
XX
PF      21-JUN-1999; 99WO-GB01949.
XX
PR      19-JUN-1998; 98GB-0013345.
XX
PA      (BIOG-) BIOGEMMA UK LTD.
XX
PI      Draper J, Kenton P, Paul W;
XX
DR      WPI, 2000-106107/09.
XX
PT      Novel promoters used to control the expression of heterologous genes in
PT      transformed plants
XX
PS      Example 3; Page 32; 67pp; English.
XX
CC      The present DNA sequence is a PCR primer-2, used for the isolation of
CC      AopRT-L promoter region. The promoter sequence was obtained from pPCR-TA
CC      using this primer. This primer is designed against both the 5' and 3' ends
CC      of the promoter, with extensions to provide appropriate restriction
CC      sites for further cloning.
XX
SQ      Sequence 41 BP; 4 A; 8 C; 14 G; 15 T; 0 other;
Query Match      4.2%; Score 20; DB 21; Length 41;
Best Local Similarity 100.0%; Pred. No. 3,3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      453 AAAACAACACACCAATC 472
      |||||||
Db      41 AAAACAACACACCAATC 22

-RESULT 10
AA229520
ID      AA229520 standard; DNA; 44 BP.
XX
AC      AA229520;
XX
DT      14-MAR-2000 (first entry)
XX
DR      Primer-1 used for construction of AopRT-Lx3 promoter.
```

```
XX      Inducible promoter; Thaumatin-like PR-5 related gene; AopRT-L; primer;
KM      non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KM      development; GUS construct; multimerisation; SA responsive element;
KM      systemic activation; AopRT-Lx3 promoter; p22-JIT60; pUC19; ss.
OS      Synthetic.
XX
PN      WO9966057-A2.
XX
PD      23-DEC-1999.
XX
PF      21-JUN-1999; 99WO-GB01949.
XX
PR      19-JUN-1998; 98GB-0013345.
XX
PA      (BIOG-) BIOGEMMA UK LTD.
XX
PI      Draper J, Kenton P, Paul W;
XX
DR      WPI, 2000-106107/09.
XX
PT      Novel promoters used to control the expression of heterologous genes in
PT      transformed plants
XX
PS      Example 12; Page 41; 67pp; English.
XX
CC      The present DNA sequence is a PCR primer-1, used for the construction of
CC      the AopRT-Lx3 promoter. This primer is used to amplify the region from
CC      0 bp to -247 bp of the AopRT-L promoter, from p22-JIT60 and cloned into
CC      pUC19.
XX
SQ      Sequence 44 BP; 10 A; 11 C; 10 G; 13 T; 0 other;
Query Match      4.0%; Score 19; DB 21; Length 44;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      226 TTGGAACTGATACCTAC 244
      |||||||
Db      26 TTGGAACTGATACCTAC 44

RESULT 11
AA229522
ID      AA229522 standard; DNA; 44 BP.
XX
AC      AA229522;
XX
DT      14-MAR-2000 (first entry)
XX
DE      Primer-3 used for construction of AopRT-Lx3 promoter.
XX
KM      Inducible promoter; Thaumatin-like PR-5 related gene; AopRT-L; primer;
KM      non-phytoxic inducing agent; Salicylic acid; SA; BTH; environmental;
KM      development; GUS construct; multimerisation; SA responsive element;
KM      systemic activation; AopRT-Lx3 promoter; p22-JIT60; pUC19; ss.
OS      Synthetic.
XX
PN      WO9966057-A2.
XX
PD      23-DEC-1999.
XX
PF      21-JUN-1999; 99WO-GB01949.
XX
PR      19-JUN-1998; 98GB-0013345.
XX
PA      (BIOG-) BIOGEMMA UK LTD.
XX
PI      Draper J, Kenton P, Paul W;
XX
DR      WPI, 2000-106107/09.
```

XX Novel promoters used to control the expression of heterologous genes in  
PT transformed plants -  
PT  
XX  
PS "Example 12: Page 41; 67pp: English.  
XX  
CC The present DNA sequence is a PCR primer-3, used for the construction of  
CC the AOPR1-Lx3 promoter. This primer is used to amplify the region from  
CC -133 bp to -247 bp of the AOPR1-L promoter, from p22-JIT60 and cloned  
CC into pUC19.  
XX  
SQ Sequence 44 BP; 10 A; 11 C; 10 G; 13 T; 0 other:  
Query Match 4.0%; Score 19; DB 21; Length 44;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 226 TTGGAACGTGAATACCTAC 244  
Db 26 TTGGAACGTGAATACCTAC 44  
RESULT 12  
AAH29638/c  
ID AAH29638 standard; DNA: 47 BP.  
XX  
AC AAH29638:  
XX  
DE 17-JUL-2001 (first entry)  
XX  
XX Drosophila melanogaster essential gene fragment, SEQ ID NO: 827.  
DE  
XX Drosophila melanogaster; fruit fly; essential gene; screening assay:  
KM pesticide; crop protection; chromosome 2; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200118547-A1.  
XX  
PD 15-MAR-2001.  
XX  
PE 06-SEP-2000; 2000MO-GB03444.  
XX  
PR 07-SEP-1999; 99GB-0021009.  
XX  
PA (UNIU ) UNIV GLASGOW.  
XX  
PI Davies RW, Kaiser K, Yang MY:  
DR WPI; 2001-281436/29.  
XX  
PT Screening assays for used for identifying compounds having a  
PT physiological effect on proteins identified as being essential -  
XX  
PS Claim 1; Page 642; 695pp; English.  
XX  
CC The present sequence is part of an essential gene from Drosophila  
CC melanogaster. Lack of expression of the protein encoded by this  
CC gene leads to a lethal or semi-lethal phenotype. The invention  
CC relates to 902 nucleic acid sequences from genes encoding proteins  
CC which are thought to be essential, and to a screening assay for  
CC identifying compounds which have a physiological effect on these  
CC proteins. Suitable compounds are useful as pesticides and may be used  
CC in conjunction with other pesticides and herbicides for crop  
CC protection. The gene corresponding to the present sequence is located  
CC on chromosome 2.  
XX  
SQ Sequence 47 BP; 13 A; 6 C; 9 G; 19 T; 0 other:  
Query Match 4.0%; Score 19; DB 22; Length 47;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 ATTACAACCTTTGCAATA 267  
Db 41 ATTACAACCTTTGCAATA 23  
RESULT 13  
AAC44332/c  
ID AAC44332 standard; DNA: 507 BP.  
XX  
AC AAC44332;  
XX  
DE 18-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 42470.  
DE  
XX Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127467.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.



ID AAF63749 strand; DNA; 1143 BP.  
XX AAF63749;  
AC  
XX  
-DT 03-APR-2001 (first entry)  
XX  
XX Drosophila gustatory receptor GR39D.2c DNA sequence.  
DE  
XX Gustatory receptor; fruit fly; taste; pheromone; semiochemical;  
XX crop damage; pest control; ds.  
OS  
XX Drosophila melanogaster.  
PN WO200077208-A2.  
XX  
XX 21-DEC-2000.  
PD  
XX 14-JUN-2000; 2000WO-US16211.  
XX  
XX 14-JUN-1999; 99US-0138668.  
PR 10-FEB-2000; 2000US-0181704.  
XX  
XX (UYVA ) UNIV YALE.  
XX  
XX Carlson PJ, Clyne PJ, Warr CG;  
PI WPI; 2001-061873/07.  
DR P-PSDB; AAB75210.  
XX  
XX New isolated nucleic acid molecule encoding Drosophila gustatory  
PT Receptor protein useful for e.g. identification of compounds which may  
PT be used for pest management.  
XX  
XX Claim 3, Page 134-135; 227pp; English.  
PS  
XX This invention relates to polynucleotide sequences AAF63732 - AAF63777  
CC which encode Drosophila gustatory receptor proteins represented by  
CC sequences AAB75193 - AAB75238. The invention includes methods for  
CC determining gustatory receptor ligands. Also included is a method for  
CC modulating the expression of the DNA encoding the receptors. The DNA and  
CC protein sequences may be used for the identification of compounds,  
CC e.g. pheromones and other semiochemicals, which may be used for pest  
CC management. The DNA sequences may also be used for behavioural studies  
CC involving gustatory systems in various organisms. Also, the DNA sequences  
CC may also be used to track down gustatory receptor genes in insects that  
CC damage crops or transmit diseases.  
XX  
XX Sequence 1143 BP; 326 A; 232 C; 226 G; 359 T; 0 other;  
SQ  
Query Match 4.0%; Score 19; DB 22; Length 1143;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 249 ATTCAACTTTGCAATA 267  
DB 641 ATTCAACTTTGCAATA 659  
RESULT 15  
AAC36210/C  
ID AAC36210 standard; DNA; 1161 BP.  
XX  
XX AAC36210;  
AC  
XX 17-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 12948.  
DE  
XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
XX Arabidopsis thaliana.

XX  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
XX  
XX 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
XX  
XX 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
XX  
XX 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
XX  
XX 16-APR-1999; 99US-0128714.  
PR 19-APR-1999; 99US-0129845.  
XX  
XX 21-APR-1999; 99US-0130077.  
PR 23-APR-1999; 99US-0130449.  
XX  
XX 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
XX  
XX 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
XX  
XX 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
XX  
XX 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
XX  
XX 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
XX  
XX 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
XX  
XX 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
XX  
XX 18-MAY-1999; 99US-0134941.  
PR 19-MAY-1999; 99US-0135124.  
XX  
XX 21-MAY-1999; 99US-0135329.  
PR 24-MAY-1999; 99US-0135629.  
XX  
XX 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
XX  
XX 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137322.  
XX  
XX 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
XX  
XX 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
XX  
XX 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
XX  
XX 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
XX  
XX 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
XX  
XX 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
XX  
XX 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
XX  
XX 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
XX  
XX 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
XX  
XX 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
XX  
XX 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
XX  
XX 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
XX  
XX 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
XX  
XX 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
XX  
XX 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143624.  
 PR 14-JUL-1999; 99US-0144005.  
 PR 15-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 16-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144862.  
 PR 20-JUL-1999; 99US-0144864.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145152.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145813.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147152.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154479.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0158293.  
 PR 13-OCT-1999; 99US-0158294.  
 PR 13-OCT-1999; 99US-0158295.  
 PR 14-OCT-1999; 99US-0158329.  
 PR 14-OCT-1999; 99US-0158330.  
 PR 14-OCT-1999; 99US-0158331.  
 PR 14-OCT-1999; 99US-0158637.  
 PR 14-OCT-1999; 99US-0158638.  
 PR 18-OCT-1999; 99US-0158584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 4.0%; Score 19; DB 21; Length 1161;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 11 TCGGACCTGACTCTCTGT 29  
 Db 370 TCGGACCTGACTCTCTGT 352

Search completed: July 20, 2003, 06:25:08  
 Job time : 190 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:21:55 ; Search time 43 Seconds  
(without alignments)  
3387.708 Million cell updates/sec

Title: US-09-719-002-1

Sequence: 1 gaattctattgcgaactga.....acaacacacacacatcatg 475

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_NA:  
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2.6/ptodata/1/ina/PCRTUS.COMB.seq:\*  
6: /cgn2.6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	3.6	447	US-09-134-001C-803	Sequence 803, App
2	17	3.6	1767	US-09-347-594-1	Sequence 1, Appl
3	17	3.6	1767	US-09-347-594-2	Sequence 2, Appl
4	17	3.6	1767	US-09-082-558-1	Sequence 1, Appl
5	17	3.6	1767	US-09-082-558-2	Sequence 2, Appl
6	17	3.6	1767	US-09-161-092-1	Sequence 1, Appl
7	17	3.6	1767	US-09-161-092-2	Sequence 2, Appl
8	17	3.6	1767	US-09-347-594-3	Sequence 3, Appl
9	17	3.6	1767	US-09-347-594-4	Sequence 4, Appl
10	17	3.6	1767	US-09-082-558-3	Sequence 3, Appl
11	17	3.6	1767	US-09-082-558-4	Sequence 4, Appl
12	17	3.6	1767	US-09-082-558-6	Sequence 6, Appl
13	17	3.6	1767	US-09-161-092-3	Sequence 3, Appl
14	17	3.6	1767	US-09-161-092-4	Sequence 4, Appl
15	17	3.6	1767	US-09-161-092-6	Sequence 6, Appl
16	17	3.6	2281	US-08-164-614A-1	Sequence 1, Appl
17	17	3.6	2281	US-08-456-4898-1	Sequence 1, Appl
18	17	3.6	2281	PCT-US93-01720-1	Sequence 1, Appl
19	17	3.6	3935	US-09-455-777-1	Sequence 1, Appl
20	17	3.6	3048	US-08-973-273-4	Sequence 4, Appl
21	17	3.6	55827	US-09-813-133A-3	Sequence 3, Appl
22	16	3.4	431	US-09-020-956-169	Sequence 169, App
23	16	3.4	431	US-09-030-607-169	Sequence 169, App
24	16	3.4	431	US-09-605-785-169	Sequence 169, App
25	16	3.4	431	US-09-439-313-169	Sequence 169, App
26	16	3.4	431	US-09-352-616A-169	Sequence 169, App
27	16	3.4	431	US-09-232-149A-169	Sequence 169, App

## ALIGNMENTS

28	16	3.4	522	4	US-09-134-001C-1041	Sequence 1041, Ap
29	16	3.4	537	4	US-09-308-003-7	Sequence 7, Appl
30	16	3.4	720	4	US-09-540-824-6	Sequence 6, Appl
31	16	3.4	1620	2	US-08-874-186-37	Sequence 37, Appl
32	16	3.4	1716	1	US-08-348-792-1	Sequence 1, Appl
33	16	3.4	1716	2	US-08-462-738-1	Sequence 1, Appl
34	16	3.4	1716	4	US-09-199-955-1	Sequence 1, Appl
35	16	3.4	1716	4	US-08-880-875-1	Sequence 1, Appl
36	16	3.4	1761	4	US-09-134-001C-1619	Sequence 1619, Ap
37	16	3.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
38	16	3.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
39	16	3.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
40	15	3.2	51	1	US-07-977-696C-63	Sequence 63, Appl
41	15	3.2	51	1	US-08-129-930B-63	Sequence 63, Appl
42	15	3.2	51	1	US-08-976-288A-63	Sequence 63, Appl
43	15	3.2	112	2	US-08-449-287-23	Sequence 23, Appl
44	15	3.2	218	3	US-09-035-648-14	Sequence 14, Appl
45	15	3.2	218	4	US-09-001-951-14	Sequence 14, Appl

  

RESULT 1	US-09-134-001C-803
Sequence 803, Application US/09134001C	
Patent No. 6380370	
GENERAL INFORMATION:	
APPLICANT: Lynn Doucette-Stamm et al	
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC	
FILE REFERENCE: STC-007	
CURRENT APPLICATION NUMBER: US/09/134,001C	
CURRENT FILING DATE: 1998-08-13	
PRIOR APPLICATION NUMBER: US 60/064,964	
PRIOR FILING DATE: 1997-11-08	
PRIOR APPLICATION NUMBER: US 60/055,779	
PRIOR FILING DATE: 1997-08-14	
NUMBER OF SEQ ID NOS: 5674	
SEQ ID NO 803	
LENGTH: 447	
TYPE: DNA	
ORGANISM: Staphylococcus epidermidis	
US-09-134-001C-803	

  

Query Match	3.6%; Score 17; DB 4; Length 447;
Best Local Similarity	100.0%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	

  

QY	264 AATTAATATTAAGAA 280
Db	255 AATTAATATTAAGAA 271

  

RESULT 2	US-09-347-594-1/C
Sequence 1, Application US/09347594	
Patent No. 6217883	
GENERAL INFORMATION:	
APPLICANT: ALAN, Gordon M.	
APPLICANT: MESHAN, Brian M.	
APPLICANT: ELLIS, John A.	
APPLICANT: KRAKOWKA, George S.	
APPLICANT: AUDONNET, Jean-Christophe F.	
TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE	
FILE REFERENCE: 454313-2338	
CURRENT APPLICATION NUMBER: US/09/347,594	
CURRENT FILING DATE: 1999-07-01	
EARLIER APPLICATION NUMBER: 98 08777	
EARLIER FILING DATE: 1998-07-06	
NUMBER OF SEQ ID NOS: 5	
SOFTWARE: PatentIn Ver. 2.1	
SEQ ID NO 1	

```

; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-347-594-1

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 31 GTGTCGCCGAGGTGCTG 47
Db 385 GTGTCGCCGAGGTGCTG 369

RESULT 3
US-09-347-594-2/c
; Sequence 2, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRAKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/347,594
; EARLIER FILING DATE: 1999-07-01
; EARLIER APPLICATION NUMBER: 98 08777
; EARLIER FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-347-594-2

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 31 GTGTCGCCGAGGTGCTG 47
Db 385 GTGTCGCCGAGGTGCTG 369

RESULT 4
US-09-082-558-1/c
; Sequence 1, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHAREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/082,558A
; EARLIER FILING DATE: 1998-05-21
; EARLIER APPLICATION NUMBER: FR 9800873
; EARLIER FILING DATE: 1998-01-22
; EARLIER APPLICATION NUMBER: FR 9803707
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: FR 97/12382
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
```

```

; SEQ ID NO 1
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-082-558-1

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 31 GTGTCGCCGAGGTGCTG 47
Db 385 GTGTCGCCGAGGTGCTG 369

RESULT 5
US-09-082-558-2/c
; Sequence 2, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHAREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/082,558A
; EARLIER FILING DATE: 1998-05-21
; EARLIER APPLICATION NUMBER: FR 9800873
; EARLIER FILING DATE: 1998-01-22
; EARLIER APPLICATION NUMBER: FR 9803707
; EARLIER FILING DATE: 1998-03-20
; EARLIER APPLICATION NUMBER: FR 97/12382
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-082-558-2

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 31 GTGTCGCCGAGGTGCTG 47
Db 385 GTGTCGCCGAGGTGCTG 369

RESULT 6
US-09-161-092-1/c
; Sequence 1, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHAREYRE, Catherine E.
; APPLICANT: CHAPUIS, Gilles E.
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/161,092
```



```

; CURRENT FILING DATE: 1998-09-25      APPLICATION NUMBER: 09/082,558
; PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-05-21
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-161-092-1

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369

RESULT 7
US-09-161-092-2/C
; Sequence 2, Application US/09161092
; Patent No. 6391314
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
; APPLICANT: CHARREYRE, Catherine E.
; APPLICANT: CHAPPUIS, Gilles E.
; TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
; TITLE OF INVENTION: REAGENTS
; FILE REFERENCE: ALLAN
; CURRENT APPLICATION NUMBER: US/09/161,092
; CURRENT FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/082,558
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9800873
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 9803707
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: FR 97/12382
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1767
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-161-092-2

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369

RESULT 8
US-09-347-594-3/C
; Sequence 3, Application US/09347594
; Patent No. 6217883
```

```

; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRACKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/347,594
; CURRENT FILING DATE: 1999-07-01
; EARLIER FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-347-594-3

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369

RESULT 9
US-09-347-594-4/C
; Sequence 4, Application US/09347594
; Patent No. 6217883
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon M.
; APPLICANT: MEEHAN, Brian M.
; APPLICANT: ELLIS, John A.
; APPLICANT: KRACKOWKA, George S.
; APPLICANT: AUDONNET, Jean-Christophe F.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS AND PARVOVIRUS VACCINE
; FILE REFERENCE: 454313-2338
; CURRENT APPLICATION NUMBER: US/09/347,594
; CURRENT FILING DATE: 1999-07-01
; EARLIER FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Porcine circovirus
US-09-347-594-4

Query Match
Best Local Similarity 100.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369

RESULT 10
US-09-082-558-3/C
; Sequence 3, Application US/09082558A
; Patent No. 6368601
; GENERAL INFORMATION:
; APPLICANT: ALLAN, Gordon
; APPLICANT: MEEHAN, Brian
; APPLICANT: CLARK, Edward
; APPLICANT: HAINES, Deborah
; APPLICANT: HASSARD, Lori
; APPLICANT: HARDING, John
```

```
APPLICANT: CHAREYRE, Catherine E.
APPLICANT: CHAPPUIS, Gilles E.
TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
FILE OF INVENTION: REAGENTS
FILE REFERENCE: ALLAN
CURRENT APPLICATION NUMBER: US/09/082,558A
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: FR 9800873
EARLIER FILING DATE: 1998-01-22
EARLIER APPLICATION NUMBER: FR 9803707
EARLIER FILING DATE: 1998-03-20
EARLIER APPLICATION NUMBER: FR 97/12382
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1768
TYPE: DNA
ORGANISM: Porcine circovirus
US-09-082-558-3
```

```
Query Match
Best Local Similarity 3.6%; Score 17; DB 4; Length 1768;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369
```

```
RESULT 11
US-09-082-558-4/c
Sequence 4, Application US/09082558A
Patent No. 6368601
GENERAL INFORMATION:
```

```
APPLICANT: ALLAN, Gordon
APPLICANT: MEEHAN, Brian
APPLICANT: CLARK, Edward
APPLICANT: HAINES, Deborah
APPLICANT: HASSARD, Lori
APPLICANT: HARDING, John
APPLICANT: CHAREYRE, Catherine E.
APPLICANT: CHAPPUIS, Gilles E.
TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
FILE OF INVENTION: REAGENTS
FILE REFERENCE: ALLAN
CURRENT APPLICATION NUMBER: US/09/082,558A
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: FR 9800873
EARLIER FILING DATE: 1998-01-22
EARLIER APPLICATION NUMBER: FR 9803707
EARLIER FILING DATE: 1998-03-20
EARLIER APPLICATION NUMBER: FR 97/12382
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1768
TYPE: DNA
ORGANISM: Porcine circovirus
US-09-082-558-4
```

```
Query Match
Best Local Similarity 3.6%; Score 17; DB 4; Length 1768;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 31 GTGCTGCCGAGGTGCTG 47
DB 385 GTGCTGCCGAGGTGCTG 369
```

```
RESULT 12
US-09-082-558-6/c
```

```
Sequence 6, Application US/09082558A
Patent No. 6368601
GENERAL INFORMATION:
APPLICANT: ALLAN, Gordon
APPLICANT: MEEHAN, Brian
APPLICANT: CLARK, Edward
APPLICANT: HAINES, Deborah
APPLICANT: HASSARD, Lori
APPLICANT: HARDING, John
APPLICANT: CHAREYRE, Catherine E.
APPLICANT: CHAPPUIS, Gilles E.
TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
FILE OF INVENTION: REAGENTS
FILE REFERENCE: ALLAN
CURRENT APPLICATION NUMBER: US/09/082,558A
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: FR 9800873
EARLIER FILING DATE: 1998-01-22
EARLIER APPLICATION NUMBER: FR 9803707
EARLIER FILING DATE: 1998-03-20
EARLIER APPLICATION NUMBER: FR 97/12382
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 1768
TYPE: DNA
ORGANISM: Porcine circovirus
FEATURE:
NAME/KEY: variation
LOCATION: (1)..(1768)
OTHER INFORMATION: N represents A or C or G or T
US-09-082-558-6
```

```
Query Match
Best Local Similarity 3.6%; Score 17; DB 4; Length 1768;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 31 GTGCTGCCGAGGTGCTG 47
DB 386 GTGCTGCCGAGGTGCTG 370
```

```
RESULT 13
US-09-161-092-3/c
Sequence 3, Application US/09161092
Patent No. 6391314
GENERAL INFORMATION:
APPLICANT: ALLAN, Gordon
APPLICANT: MEEHAN, Brian
APPLICANT: CLARK, Edward
APPLICANT: HAINES, Deborah
APPLICANT: HASSARD, Lori
APPLICANT: HARDING, John
APPLICANT: CHAREYRE, Catherine E.
APPLICANT: CHAPPUIS, Gilles E.
TITLE OF INVENTION: NEW PORCINE CIRCOVIRUSES, VACCINES AND DIAGNOSTIC
FILE OF INVENTION: REAGENTS
FILE REFERENCE: ALLAN
CURRENT APPLICATION NUMBER: US/09/161,092
EARLIER FILING DATE: 1998-09-25
EARLIER APPLICATION NUMBER: 09/082,558
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: FR 9800873
EARLIER FILING DATE: 1998-01-22
EARLIER APPLICATION NUMBER: FR 9803707
EARLIER FILING DATE: 1998-03-20
EARLIER APPLICATION NUMBER: FR 97/12382
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1768
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:51:05 ; Search time 172 Seconds  
(without alignments)  
5697.240 Million cell updates/sec

Title: US-09-719-002-1

Perfect score: 475  
Sequence: 1 gaattctattgcgacctga.....acaacaacaacatcatcgt 475

Scoring table: ORIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1439767 seqs, 1031500376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq3.\*  
13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
1	21	4.4	716	15	US-10-027-632-34117	Sequence 34117, A
2	21	4.4	716	15	US-10-027-632-34118	Sequence 34118, A
3	21	4.4	1691139	15	US-10-067-514-1	Sequence 1, App11
4	19	4.0	584	15	US-10-027-632-243175	Sequence 243175, A
5	19	4.0	649	15	US-10-027-632-134120	Sequence 134120, A
6	19	4.0	649	15	US-10-027-632-134121	Sequence 134121, A
7	18	3.8	397	15	US-09-817-800A-897	Sequence 897, App
8	18	3.8	475	12	US-09-918-955-25671	Sequence 25671, A
9	18	3.8	491	15	US-10-027-632-48705	Sequence 48705, A
10	18	3.8	612	15	US-10-027-632-48706	Sequence 48706, A
11	18	3.8	612	15	US-10-027-632-61505	Sequence 61505, A
12	18	3.8	612	15	US-10-027-632-62397	Sequence 62397, A
13	18	3.8	612	15	US-10-027-632-62399	Sequence 62399, A
14	18	3.8	656	15	US-10-027-632-111366	Sequence 111366, A
15	18	3.8	694	15	US-10-027-632-48227	Sequence 48227, A
16	18	3.8	745	15	US-10-027-632-28109	Sequence 28109, A

17	18	3.8	745	15	US-10-027-632-139025	Sequence 139025, A
18	18	3.8	1492	10	US-09-822-849A-141	Sequence 141, App
19	18	3.8	1609	15	US-10-082-830-48	Sequence 48, App1
20	18	3.8	148567	11	US-09-801-876B-3	Sequence 3, App1
21	18	3.8	148567	15	US-10-254-869-3	Sequence 3, App1
22	18	3.8	1691139	15	US-10-067-514-1	Sequence 1, App1
23	17	3.6	240	11	US-09-935-428A-12	Sequence 12, App1
24	17	3.6	392	12	US-09-811-284-37	Sequence 37, App1
25	17	3.6	399	12	US-09-918-955-36916	Sequence 36916, A
26	17	3.6	410	11	US-09-960-352-3865	Sequence 3865, App
27	17	3.6	412	11	US-09-998-558-1114	Sequence 1114, App
28	17	3.6	445	12	US-09-918-955-1630	Sequence 1630, App
29	17	3.6	448	15	US-10-050-704-45	Sequence 45, App1
30	17	3.6	452	15	US-10-027-632-2244	Sequence 2244, App
31	17	3.6	454	15	US-10-027-632-108770	Sequence 108770, App
32	17	3.6	466	11	US-09-998-598-1522	Sequence 1522, App
33	17	3.6	482	11	US-09-736-457-843	Sequence 843, App
34	17	3.6	482	11	US-09-907-941-843	Sequence 843, App
35	17	3.6	482	11	US-09-849-626-843	Sequence 843, App
36	17	3.6	482	15	US-10-017-754-843	Sequence 843, App
37	17	3.6	489	12	US-09-918-955-33867	Sequence 33867, A
38	17	3.6	494	10	US-09-811-284-40	Sequence 40, App1
39	17	3.6	501	15	US-10-102-524-609	Sequence 609, App
40	17	3.6	506	11	US-09-764-847-463	Sequence 463, App
41	17	3.6	506	15	US-10-097-154-463	Sequence 463, App
42	17	3.6	541	15	US-10-027-632-321125	Sequence 321125, App
43	17	3.6	581	10	US-09-799-777-132	Sequence 132, App
44	17	3.6	634	10	US-10-027-632-252063	Sequence 252063, App
45	17	3.6	645	15	US-10-027-632-274868	Sequence 274868, App

## ALIGNMENTS

RESULT 1  
US-10-027-632-34117  
Sequence 34117, Application US/10027632  
GENERAL INFORMATION:  
APPLICANT: Mary, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108627.129  
CURRENT APPLICATION NUMBER: US/10/027, 632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218, 006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198, 676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193, 483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185, 218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167, 363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156, 358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146, 002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 34117  
LENGTH: 716  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-34117  
Query Match 4.4%; Score 21; DB 15; Length 716;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
241 CTACATTAAATACAACTTTG 261  
|||||  
361 CTACATTAAATACAACTTTG 381

```
RESULT 2
US-10-027-632-34118
; Sequence 34118, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34118
; LENGTH: 716
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-34118

Query Match
Best Local Similarity 100.0%; Score 21; DB 15; Length 716;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 CTACATTAATTACAACTTTG 261
DB 361 CTACATTAATTACAACTTTG 381

RESULT 3
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretaardt, Sif
; APPLICANT: Reynolds, Sif
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

Query Match
Best Local Similarity 100.0%; Score 21; DB 15; Length 1691139;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 4
US-10-027-632-243175/C
; Sequence 243175, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 243175
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)-(584)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-243175

Query Match
Best Local Similarity 100.0%; Score 19; DB 15; Length 584;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 AACCTTTCGAATTAATA 272
DB 156 AACCTTTCGAATTAATA 138

RESULT 5
US-10-027-632-134120
; Sequence 134120, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 134120
```

LENGTH: 649  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-134120

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 15; Length 649;  
 Pred. No. 9.8;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 418 CTGCCCATTCCTCCCTCC 436  
 |||||  
 DB 433 CTGCCCATTCCTCCCTCC 451

RESULT 6  
 US-10-027-632-134121  
 Sequence 134121, Application US/10027632  
 GENERAL INFORMATION:  
 APPLICANT: Wang, David G.  
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 TITLE OF INVENTION: Polymorphisms in the Human Genome  
 FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30  
 PRIOR APPLICATION NUMBER: US 60/218,006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198,676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193,483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185,218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167,363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156,358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146,002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 134121  
 LENGTH: 649  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-134121

Query Match  
 Best Local Similarity 100.0%; Score 19; DB 15; Length 649;  
 Pred. No. 9.8;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 418 CTGCCCATTCCTCCCTCC 436  
 |||||  
 DB 433 CTGCCCATTCCTCCCTCC 451

RESULT 7  
 US-09-917-800A-897  
 Sequence 897, Application US/09917800A  
 Patent No. US20020119462A1  
 GENERAL INFORMATION:  
 APPLICANT: Mendrick, Donna  
 APPLICANT: Porter, Mark  
 APPLICANT: Johnson, Kory  
 APPLICANT: Castle, Arthur  
 APPLICANT: Elashoff, Michael

FILE REFERENCE: 44921-5038-US  
 CURRENT APPLICATION NUMBER: US/09/917,800A  
 CURRENT FILING DATE: 2001-07-31  
 PRIOR APPLICATION NUMBER: US 60/222,040  
 PRIOR FILING DATE: 2000-07-31  
 PRIOR APPLICATION NUMBER: US 60/222,880

PRIOR FILING DATE: 2000-11-02  
 PRIOR APPLICATION NUMBER: US 60/290,029  
 PRIOR FILING DATE: 2001-05-11  
 PRIOR APPLICATION NUMBER: US 60/290,645  
 PRIOR FILING DATE: 2001-05-15  
 PRIOR APPLICATION NUMBER: US 60/292,336  
 PRIOR FILING DATE: 2001-05-22  
 PRIOR APPLICATION NUMBER: US 60/295,798  
 PRIOR FILING DATE: 2001-06-06  
 PRIOR APPLICATION NUMBER: US 60/297,457  
 PRIOR FILING DATE: 2001-06-13  
 PRIOR APPLICATION NUMBER: US 60/298,884  
 PRIOR FILING DATE: 2001-06-19  
 PRIOR APPLICATION NUMBER: US 60/303,459  
 PRIOR FILING DATE: 2001-07-09  
 NUMBER OF SEQ ID NOS: 1740  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 897  
 LENGTH: 397  
 TYPE: DNA  
 ORGANISM: Rattus norvegicus  
 FEATURE:  
 OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1171319  
 US-09-917-800A-897

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 11; Length 397;  
 Pred. No. 32;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 451 CAAACACACACACACAC 468  
 |||||  
 DB 55 CAAACACACACACACAC 72

RESULT 8  
 US-09-918-995-25871/C  
 Sequence 25871, Application US/09918995  
 Publication No. US20030073623A1  
 GENERAL INFORMATION:  
 APPLICANT: Hyseq, Inc.  
 TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
 TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
 FILE REFERENCE: 20411-756  
 CURRENT APPLICATION NUMBER: US/09/918,995  
 CURRENT FILING DATE: 2001-07-30  
 PRIOR APPLICATION NUMBER: US/09/235,076  
 PRIOR FILING DATE: 1999-01-20  
 NUMBER OF SEQ ID NOS: 38054  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 25871  
 LENGTH: 475  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)..(475)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-918-995-25871

Query Match  
 Best Local Similarity 100.0%; Score 18; DB 12; Length 475;  
 Pred. No. 32;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 263 AATATAATATAAGAAA 280  
 |||||  
 DB 68 AATATAATATAAGAAA 51

RESULT 9  
 US-10-027-632-48705  
 Sequence 48705, Application US/10027632  
 GENERAL INFORMATION:  
 APPLICANT: Wang, David G.

```

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48705
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48705
```

```

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 491;
Matches 18; Conservative 100.0%; Pred. No. 32; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 100 CAATATATTTCATCT 117
DB 439 CAATATATTTCATCT 456
```

```

RESULT 10
US-10-027-632-48706
; Sequence 48706, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/228,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48706
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48706
```

```

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 491;
Matches 18; Conservative 100.0%; Pred. No. 32; Mismatches 0; Indels 0; Gaps 0;
OY 100 CAATATATTTCATCT 117
```

```

DB 439 CAATATATTTCATCT 456
```

```

RESULT 11
US-10-027-632-61505
; Sequence 61505, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61505
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-61505
```

```

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 612;
Matches 18; Conservative 100.0%; Pred. No. 33; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 263 AATTAATATTAAGAA 280
DB 402 AATTAATATTAAGAA 419
```

```

RESULT 12
US-10-027-632-62397
; Sequence 62397, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62397
; LENGTH: 612
```

```

; TYPE: DNA
; ORGANISM: Human
US-10-027-632-62397

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 612;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AATATAATATAAGAAA 280
    |||||||
Db 402 AATATAATATAAGAAA 419

RESULT 13
US-10-027-632-62329
; Sequence 63229, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63229
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-63229

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 612;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AATATAATATAAGAAA 280
    |||||||
Db 402 AATATAATATAAGAAA 419

RESULT 14
US-10-027-632-111366/C
; Sequence 111366, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
```

```

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111366
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111366

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 656;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 444 TCTACTCAAAACACACA 461
    |||||||
Db 385 TCTACTCAAAACACACA 368

RESULT 15
US-10-027-632-48227
; Sequence 48227, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48227
; LENGTH: 694
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-48227

Query Match
Best Local Similarity 3.8%; Score 18; DB 15; Length 694;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AATATAATATAAGAAA 280
    |||||||
Db 392 AATATAATATAAGAAA 409

Search completed: July 20, 2003, 07:53:40
Job time : 176 secs
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: July 20, 2003, 06:08:35 ; Search time 1127 Seconds

(without alignments)  
6825.959 Million cell updates/sec

Title: US-09-719-002-1

Perfect score: 475  
Sequence: 1 gaattcttattgcgacctga.....acaacacacacacacatcatc 475

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:  
1: em\_estb1:  
2: em\_estb2:  
3: em\_estb3:  
4: em\_estb4:  
5: em\_estb5:  
6: em\_estb6:  
7: em\_estb7:  
8: em\_estb8:  
9: em\_estb9:  
10: em\_estb10:  
11: em\_estb11:  
12: em\_estb12:  
13: em\_estb13:  
14: em\_estb14:  
15: em\_estb15:  
16: em\_estb16:  
17: em\_estb17:  
18: em\_estb18:  
19: em\_estb19:  
20: em\_estb20:  
21: em\_estb21:  
22: em\_estb22:  
23: em\_estb23:  
24: em\_estb24:  
25: em\_estb25:  
26: em\_estb26:  
27: em\_estb27:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	21	4.4	252	10 BE370506 BE370506 601218294
2	21	4.4	426	10 BE147853 BE147853 RC3-HT023
3	21	4.4	587	10 BE205326 BE205326 EST398002
4	21	4.4	669	13 BE311957 BE311957 EST51370
5	21	4.4	769	17 AG143540 AG143540 Pan trogl
6	21	4.4	822	17 BH371846 BH371846 AG-ND-162

7	21	4.4	845	12 BE573809 BE573809 602132006
8	21	4.4	970	17 CENS030N2 AL2222455 Tetradon
9	20	4.2	401	10 BB818014 BB818014
10	20	4.2	401	13 BT790095 BT790095
11	20	4.2	405	9 AA461949 AA461949
12	20	4.2	408	10 BB810434 BB810434
13	20	4.2	413	17 A327901 A327901
14	20	4.2	418	12 BE225508 BE225508
15	20	4.2	420	13 BG975401 BG975401
16	20	4.2	430	12 BG143281 BG143281
17	20	4.2	432	9 AU043619 AU043619
18	20	4.2	449	13 BM502635 BM502635
19	20	4.2	468	12 BG071544 BG071544
20	20	4.2	472	9 A1227421 A1227421
21	20	4.2	475	9 AA615561 AA615561
22	20	4.2	475	10 BE448225 BE448225
23	20	4.2	495	10 AM214359 AM214359
24	20	4.2	496	17 A2767485 A2767485
25	20	4.2	512	10 AM214512 AM214512
26	20	4.2	518	13 BM502321 BM502321
27	20	4.2	532	9 AU019722 AU019722
28	20	4.2	538	10 BE631935 BE631935
29	20	4.2	542	13 BM222085 BM222085
30	20	4.2	547	9 AU040283 AU040283
31	20	4.2	552	13 BM229851 BM229851
32	20	4.2	559	9 AU021058 AU021058
33	20	4.2	572	14 BM900102 BM900102
34	20	4.2	574	13 BM225839 BM225839
35	20	4.2	580	13 BM228549 BM228549
36	20	4.2	586	13 BM229869 BM229869
37	20	4.2	607	12 BG084386 BG084386
38	20	4.2	656	13 BM228341 BM228341
39	20	4.2	657	17 AZ055055 AZ055055
40	20	4.2	664	13 BM228469 BM228469
41	20	4.2	677	10 BB399841 BB399841
42	20	4.2	701	17 BB299872 BB299872
43	20	4.2	909	13 BT39973 BT39973
44	20	4.2	1058	12 BF143661 BF143661
45	19	4.0	270	14 F19925 F19925

#### ALIGNMENTS

RESULT 1  
LOCUS BE370506  
DEFINITION 601218294F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3587470 5',  
ACCESSION BE370506  
VERSION BE370506.1 GI:9315869  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 252)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM8750 row: k column: 23  
High quality sequence stop: 242.  
Location/Qualifiers  
1. 252

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:3587470"  
 /clone\_lib="NCI-CGAP-Lu29"  
 /tissue\_type="Spontaneous tumor, metastatic to mammary.  
 Stem cell origin."  
 /lab\_host="Dh10B"  
 /note="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;  
 Site: 2: NotI. Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 75 a 61 c 43 g 73 t  
 ORIGIN

Query Match 4.4%; Score 21; DB 10; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 255 ACCTTGCATAAATAATA 275  
 DB 232 ACCTTGCATAAATAATA 252

RESULT 2 BE147853 426 bp mRNA linear EST 21-JUN-2000  
 LOCUS RC3-HR0230-201199-013-g05 HR0230 Homo sapiens cDNA, mRNA sequence.  
 DEFINITION BE147853  
 ACCESSION BE147853.1 GI:8610577  
 VERSION BE147853.1 GI:8610577  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 426)  
 AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,  
 Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
 Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
 Simpson, A. J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL PROC. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL:  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=652-RC3-HR0230-201  
 199-013-g05&ts=1999-11-20&td=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 105  
 High quality sequence stop: 409.  
 Location/Qualifiers

## FEATURES

1. 426  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HR0230"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site: 1: SmaI;  
 Site: 2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under

BASE COUNT 109 a 105 c 121 g 91 t  
 ORIGIN

Query Match 4.4%; Score 21; DB 10; Length 426;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 CTCAAAACACACACACCA 469  
 DB 406 CTCAAAACACACACACCA 426

RESULT 3 BE205326 587 bp mRNA linear EST 05-SEP-2000  
 LOCUS BE205326  
 DEFINITION EST198002 KYO Medicago truncatula cDNA clone PKV0-21E10, mRNA  
 sequence.  
 ACCESSION BE205326  
 VERSION BE205326.1 GI:8748622  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 587)  
 AUTHORS Vandenbosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L.,  
 Town, C. D., Bowman, C. L., Craven, M. B., Hansen, T. S., Holt, L. E. and  
 Fraser, C. M.  
 TITLE ESTs from uninoculated seedling roots of Medicago truncatula  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Vandenbosch K  
 Department of Plant Biology  
 University of Minnesota  
 220 Biosci Center, 1443 Gortner Ave, St. Paul, MN 55108, USA  
 Tel: 612 624 2755  
 Fax: 612 625 1738  
 Email: kvandenbosch@umn.edu  
 Texas A&M University name: T265098e  
 TIGR sequence name: MTGBL29TK  
 More information is available at:  
 http://chryslr.tamu.edu/medicago  
 Seq primer: SMOd (CTA GAA CTA gtc gAT CC).  
 Location/Qualifiers

## FEATURES

1. 587  
 /organism="Medicago truncatula"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone\_lib="PKV0-21E10"  
 /clone\_lib="KY0"  
 /tissue\_type="Seedling roots"  
 /dev\_stage="Immediately prior to inoculation with  
 Sinorhizobium meliloti (0 hour)"  
 /lab\_host="E. coli strain XL0R"  
 /note="Vector: Bluescript SK-; Site: 1: EcoRI; Site: 2:  
 XhoI. cDNA was prepared from polyA+ enriched RNA. The cDNA  
 was directionally ligated into the Unizap XR vector from  
 Stratagene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-assist  
 helper phage and propagated in XL0R cells."

BASE COUNT 164 a 69 c 143 g 211 t  
 ORIGIN

Query Match 4.4%; Score 21; DB 10; Length 587;  
 Best Local Similarity 100.0%; Pred. No. 5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 454 AAACAACACACACCAATCAT 474  
 DB 54 AAACAACACACACCAATCAT 34

RESULT 4  
 B1311957/c  
 LOCUS  
 DEFINITION B1311957 669 bp mRNA linear EST 20-JUL-2001  
 EST5313707 GSSD Medicago truncatula cDNA clone pGSD16D21 5' end,  
 mRNA sequence.  
 ACCESION B1311957  
 VERSION B1311957  
 KEYWORDS GI:14986284  
 SOURCE  
 ORGANISM  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 REFERENCE  
 1 (bases 1 to 669)  
 Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T., Cho,  
 J. and Fraser, C.M.  
 ESTs from developing reproductive tissues of Medicago truncatula  
 Unpublished (2001)  
 CONTACT: Michael A. Grusak  
 USDA/ARS Children's Nutrition Research Center  
 Baylor College of Medicine  
 1100 Bates Street, Houston, TX 77030-2600, USA  
 Tel: 713-798-7044  
 Fax: 713-798-7078  
 Email: mgrusak@bcm.tmc.edu  
 B398812e  
 TIGR sequence name: MTPBE233K  
 More information is available at: www.medicago.org  
 Seq primer: SKmcd (CTA GAA CTA gtc gat CC).  
 FEATURES  
 source  
 1. 669  
 /organism="Medicago truncatula"  
 /cultivar="A17"  
 /db\_xref="taxon:3880"  
 /clone="pGSD16D21"  
 /clone\_lib="GSSD"  
 /tissue\_type="immature seeds"  
 /dev\_stage="immature seeds, 11 to 19 days after  
 pollination"  
 /note="Vector: p Bluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Immature seeds, collected from pods ranging in age  
 from 11 to 19 days after pollination, were harvested from  
 greenhouse-grown plants. Seeds were removed and  
 separated from pod walls and were immediately frozen in  
 liquid nitrogen. Seeds throughout the age range were  
 pooled for mRNA extraction. cDNA was prepared from polyA+  
 enriched RNA. The cDNA was directionally ligated into  
 the UniZap XR vector from stragene and packaged using  
 Gigapack III Gold packaging extracts. Plasmids containing  
 cDNA inserts were excised from the recombinant lambda-zap  
 phage using Ex-assist helper phage and propagated in  
 XLOLR cells."  
 BASE COUNT 178 a 86 c 182 g 223 t  
 ORIGIN  
 Query Match 4.4%; Score 21; DB 13; Length 669;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 454 AACACACACACACCATCAT 474  
 ||||||||||||||||||  
 -DB 571 AACACACACACACCATCAT 551  
 ||||||||||||||||||  
 RESULT 5  
 AG143540 769 bp DNA linear GSS 08-JAN-2002  
 LOCUS  
 DEFINITION Pan troglodytes DNA, clone: RP43-004E06.TJ, genomic survey  
 sequence.  
 ACCESSION AG143540

VERSION  
 AG143540.1 GI:16673218  
 GSS.  
 KEYWORDS  
 Pan troglodytes male lymphocytes DNA, clone\_lib:RP43-004E06.TJ.  
 SOURCE  
 Male BAC Library clone:RP43-004E06.TJ.  
 ORGANISM  
 Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
 REFERENCE  
 1  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 BAC end sequences of Library RP43-43  
 Unpublished  
 TITLE  
 2 (bases 1 to 769)  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totoki, Y., Watanabe, H. and Sakaki, Y.  
 DIRECT SUBMISSION  
 Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimpses@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp//,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library RP43-43 This BAC  
 end was generated during the R&D process and may have higher chance  
 of clone tracking errors.  
 PRIMERS  
 Sequencing: TJ  
 LIBRARY  
 Vector : pBACe3.6  
 R.Site 1 : EcoRI  
 R.Site 2 : EcoRI.  
 Location/Qualifiers  
 1. 769  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="RP43-004E06.TJ"  
 /sex="male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RP43-43 Chimpanzee Male BAC Library"  
 BASE COUNT 185 a 203 c 109 g 270 t  
 ORIGIN  
 Query Match 4.4%; Score 21; DB 17; Length 769;  
 Best Local Similarity 100.0%; Pred. No. 5.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 241 CTACATTAAATTAACAATTGG 261  
 ||||||||||||||||||  
 -DB 103 CTACATTAAATTAACAATTGG 123  
 ||||||||||||||||||  
 RESULT 6  
 BH371846 822 bp DNA linear GSS 10-DEC-2001  
 LOCUS  
 DEFINITION Ag-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17  
 DNA sequence.  
 ACCESION BH371846  
 VERSION BH371846.1 GI:17317971  
 GSS.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.  
 REFERENCE  
 1 (bases 1 to 822)  
 Shetty, V., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.  
 Direct Submission of BAC-end sequences from Anopheles gambiae  
 Unpublished (2001)  
 Other GSSs: AG-ND-162M17.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For

Class: BAC ends.

#### FEATURES

source

Location/Qualifiers

1..822

/organism="Anopheles gambiae"

/strain="PST"

/db\_xref="taxon:7165"

/clone\_id="AG-ND-162M17"

/clone\_lib="ND-TAM"

/note="Vector: pCIBAC1; Site\_1: HindIII"

#### BASE COUNT

235 a 142 c 155 g 290 t

#### ORIGIN

Query Match

Best Local Similarity 4.4%; Score 21; DB 17; Length 822;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

256 CTTTGCAATTAATATATAA 276

Db

139 CTTTGCAATTAATATATAA 159

#### RESULT 7

LOCUS

BF573809 845 bp MRNA 1linear EST 12-DEC-2000

#### DEFINITION

602132006F1 NIH-MGC-81 Homo sapiens cDNA clone IMAGE:4271376 5',

MRNA sequence.

#### ACCESSION

BF573809

#### VERSION

BF573809.1 GI:11647521

#### KEYWORDS

EST

#### SOURCE

human

#### ORGANISM

Homo sapiens

#### REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

#### AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

#### TITLE

NIH-MGC http://mgc.nci.nih.gov/.

#### JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

#### COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L10C1083 row: 1 column: 01

High quality sequence stop: 165.

#### FEATURES

source

Location/Qualifiers

1..845

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="IMAGE:4271376"

/clone\_lib="NIH-MGC-81"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: muscle (skeletal); Vector: pDNR-LIB

(Clontech); Site\_1: SfiI (ggccattagcc); 5' and 3' adaptors were used in cloning

as follows: 5' adaptor sequence: 5'-CACGGCCATATAGCC-3'

and 3' adaptor sequence: 5'-ATTCTAGAGCCGAGCGCCGACATG-dt(30)BN-3' (where B = A,

#### BASE COUNT

261 a 118 c 241 g 225 t

#### ORIGIN

Query Match

Best Local Similarity 4.4%; Score 21; DB 12; Length 845;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

337 TTCCACAAAGGCTTAGACTTT 357

Db

394 TTCCACAAAGGCTTAGACTTT 414

#### RESULT 8

CNS030N2

#### LOCUS

Tetradon nigroviridis genome survey sequence T7 end of clone

#### DEFINITION

203E01 of library G from Tetradon nigroviridis, genomic survey

#### ACCESSION

AL222455.1 GI:7881274

#### VERSION

AL222455.1 GI:7881274

#### KEYWORDS

GSS: genome survey sequence.

#### SOURCE

Tetradon nigroviridis.

#### ORGANISM

Tetradon nigroviridis.

#### REFERENCE

1 (bases 1 to 970)

#### AUTHORS

Roeest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,

#### TITLE

Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,

#### JOURNAL

Human gene number estimate provided by genome wide analysis using

#### AUTHORS

Unpublished

#### COMMENT

Submitted (12-APR-2000)

#### TITLE

This sequence is a single read and was generated as part of a large

#### AUTHORS

scale clone-end sequencing project of the tetradon nigroviridis

#### TITLE

genome. For more information, please take a look at

#### JOURNAL

http://www.genoscope.cns.fr/Tetradon.

#### COMMENT

Location/Qualifiers

#### FEATURES

1..970

#### source

/organism="Tetradon nigroviridis"

#### ORIGIN

/db\_xref="taxon:99883"

#### BASE COUNT

133 a 334 c 342 g 145 t 16 others

#### Query Match

Best Local Similarity 4.4%; Score 21; DB 17; Length 970;

#### Matches

21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

32 TGCTGCGAGGCTGCTGCGAA 52

#### Db

480 TGCTGCGAGGCTGCTGCGAA 500

RESULT 9  
B818014 401 bp mRNA linear EST 19-NOV-2001  
LOCUS B818014 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus  
DEFINITION Musculus cDNA clone G730042D17 3', mRNA sequence.  
B818014  
VERSION B818014.1 GI:16990643  
KEYWORDS EST.  
ACCESSION  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 401)  
REFERENCE  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hangaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kondo,M., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takeku-Akaira,S., Tanaka,T., Tomaru,A., Toyota,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,F., et al. 2001)  
TITLE Unpublished (2001)  
JOURNAL  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp/  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10): 1617-1630 (2000)  
vagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11): 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2): 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.  
FEATURES  
Source  
Location/Qualifiers  
1..401  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G730042D17"  
/clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"  
/tissue\_type="lung"  
/cell\_line="RCB-0558 LLC"  
/note="Pooled cell lines: (cell\_line=CRU-1751 WEHI 164), (cell\_line=CRU-2116 JC), (cell\_line=RCB-0035 WEHI-3), (cell\_line=RCB-0464 Meth-A), (cell\_line=RCB-0545 OHTA), (cell\_line=RCB-0559 K-1.F1), (cell\_line=RCB-1283 B16 melanoma), (cell\_type=B cells), (cell\_line=CRU-1702 WEHI 231), (cell\_type=Leydig cells), (cell\_line=CRU-2065 MTC-1), (cell\_type=Nullipotent stem cell), (cell\_line=CRU-2070 NE), (tissue\_type=bladder), (cell\_line=RCB-0544 MBT-2), (tissue\_type=bone marrow), (cell\_type=stroma cell), (cell\_line=CRU-2028 SR-4987), (tissue\_type=colon), (cell\_line=RCB-0549 Cle-H3), (tissue\_type=kidney),

cell\_line=CCL-142 RAG), (tissue\_type=submandibular gland, cell\_line=CRU-1734 SCA-9 clone 15), (strain=BALB/C, cell\_type=B cells), cell\_line=CRU-1669 BCL1 clone 13.20-3B3), (strain=C3H, tissue\_type=brain, cell\_line=CRU-1443 BC3H1)"  
BASE COUNT 132 a 86 c 57 g 126 t  
ORIGIN  
Query Match 4.2%; Score 20; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches: 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 413 AATACCTGCCATTCCTCC 432  
DB 14 AATACCTGCCATTCCTCC 33  
RESULT 10  
LOCUS B1790095/c  
DEFINITION Musculus cDNA clone IMAGE:5661262 3', mRNA sequence.  
B1790095 401 bp mRNA linear EST 12-MAR-2002  
1c93b12.x1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus  
musculus cDNA clone IMAGE:5661262 3', mRNA sequence.  
B1790095  
VERSION B1790095.1 GI:15917820  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 401)  
REFERENCE  
AUTHORS Melton,D., Brown,J., Kenny,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Seearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagariswill,R., Williams,T., Jackson,X. and Bowers,Y.  
Jackson,X. and Bowers,Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
TITLE  
JOURNAL  
COMMENT Other-ESTs: 1c93b12.y1  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@molb.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownj@fas.harvard.edu)  
MGI:1947588 This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov  
Seq primer: -400p from 61bco  
High quality sequence stop: 396.  
FEATURES  
Source  
Location/Qualifiers  
1..401  
/organism="Mus musculus"  
/strain="ICR"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5661262"  
/clone\_lib="Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1"  
/sex="Both for embryonic & newborn, male for adult and adult islet"  
/dev\_stage="Embryonic day 10.5, E12.5, E16.5, newborn, adult, mixed"  
/lab\_note="DH10B"  
/note="Vector: pSPORT1, Site: 1: Not I; Site: 2: Sal I; Five libraries representing E10.5/12.5 pancreatic bud, E16.5 pancreas, newborn pancreas, adult pancreas, and adult islets of langerhans were separately constructed using Superscript Plasmid library kit (Life Technologies). cDNA

was made by oligo-dT priming and size-selected by column fractionation. Libraries were amplified once on solid support and plasmid DNA from each library was prepared and mixed in equal amounts. The mixed library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806. 0.5 microgram single-stranded mixed library plasmid DNA was mixed with 5 micrograms PCR product representing mixed library inserts and hybridized to an EcoT of 6. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 127 a 56 c 85 g 133 t

Query Match 4.2% Score 20; DB 13; Length 401;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 AATACCTGCCATTCCTC 432  
|||||  
Db 389 AATACCTGCCATTCCTC 370

RESULT 11  
AA461949 405 bp mRNA linear EST 10-JUN-1997  
LOCUS Vg71c09.r1 Soares mouse NbM Mus musculus cDNA clone IMAGE:871408  
DEFINITION 5', mRNA sequence.  
ACCESSION AA461949  
VERSION AA461949.1 GI:2186840  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 405)  
Marrin, S., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schenleberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Maria M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MG1510888

Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 353.

#### FEATURES

source

1. 405  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:871408"  
/clone\_lib="Soares mouse NbM"  
/sex="male"  
/tissue\_type="heart"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGAGCGCGCGGAGAAAGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia); digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA

BASE COUNT 131 a 88 c 59 g 127 t

Query Match 4.2% Score 20; DB 9; Length 405;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 AATACCTGCCATTCCTC 432  
|||||  
Db 36 AATACCTGCCATTCCTC 55

RESULT 12  
BB810434 408 bp mRNA linear EST 19-NOV-2001  
LOCUS BB810434 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus  
DEFINITION Musculus cDNA clone G730005F13 3', mRNA sequence.  
ACCESSION BB810434  
VERSION BB810434.1 GI:16983063  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 408)  
Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hasegaki, T., Hayatsu, N., Hirokawa, K., Hirokawa, T., Imotoh, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Nomura, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyata, T., Watabiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexed capillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

#### FEATURES

source

Location/Qualifiers  
1. 408  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G730005F13"

/clone\_lib="RIKEN full-length enriched, lung RCB-0558 LLC  
 cDNA"  
 /tissue\_type="lung"  
 /cell\_line="RCB-0558 LLC"  
 /note="pooled cell lines : (cell\_line=CRJ-1751 WEHI 164),  
 (cell\_line=CRJ-2116 JC), (cell\_line=RCB-0035 WEHI-3),  
 (cell\_line=RCB-0464 Meth-A), (cell\_line=RCB-0545 OHTA),  
 (cell\_line=RCB-0559 K-1, Fl), (cell\_line=RCB-1283 B16  
 melanoma), (cell\_type=B cells, cell\_line=CRJ-1702 WEHI 231  
 ), (cell\_type=leidyg cells, cell\_line=CRJ-2065 MLTC-1),  
 (cell\_type=Nullipotent stem cell, cell\_line=CRJ-2070 NE),  
 (tissue\_type=bladder, cell\_line=RCB-0544 MBT-2),  
 (tissue\_type=bone marrow, cell\_type=stroma cell,  
 cell\_line=CRJ-2028 SR-4987), (tissue\_type=colon,  
 cell\_line=RCB-0549 Cle-M3), (tissue\_type=kidney,  
 cell\_line=CRJ-142 RAG), (tissue\_type=submandibular gland,  
 cell\_line=CRJ-1734 SCA-9 clone 15) (strain=BA1B/C,  
 cell\_type=B cells, cell\_line=CRJ-1669 BCL1 clone 13,20-383  
 ), (strain=C3H, tissue\_type=brain, cell\_line=CRJ-1443  
 BC3H)"

BASE COUNT 134 a 89 c 57 g 128 t  
 ORIGIN  
 Query Match 4.2% Score 20; DB 10; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 AATACCTGCCATTCCTC 432  
 Db 22 AATACCTGCCATTCCTC 41

RESULT 13  
 LOCUS AZ327901.c 413 bp DNA linear GSS 29-SEP-2000  
 DEFINITION 1M0051K05R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0051K05 R, DNA sequence.  
 ACCESSION AZ327901  
 VERSION AZ327901.1 GI:10387286  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 413)  
 REFERENCE  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 and Wright,D., Weiss,R., Stokes,R., Tingey,A., von Niederhausern,A.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0051 row: K column: 05  
 Seq primer: CACACAGGAAACGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 413.  
 Location/Qualifiers  
 1. 413  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_lib="UUGC1M0051K05"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: pMD29mv. Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pMD42 (g14732114bp/AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptor complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

BASE COUNT 91 a 100 c 75 g 147 t  
 ORIGIN  
 Query Match 4.2% Score 20; DB 17; Length 413;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 263 AATAAATATTAAGAAAGT 282  
 Db 113 AATAAATATTAAGAAAGT 94

RESULT 14  
 LOCUS BF225508 418 bp mRNA linear EST 29-DEC-2000.  
 DEFINITION UY41107.X1 NCI-CGAP\_Lu30 Mus musculus cDNA clone IMAGE:3662149  
 RNA sequence.  
 ACCESSION BF225508  
 VERSION BF225508.1 GI:11133070  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 418)  
 REFERENCE  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 COMMENT Other-ESTs: uy41107.y1  
 Contact: Robert Strauberg, Ph.D.  
 Email: cgaaps-r@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 image.llnl.gov/image/html/lresources.shtml  
 MGI:1422917  
 High quality sequence stop: 406.  
 Location/Qualifiers  
 1. 418  
 /organism="Mus musculus"  
 /strain="C2EHC II"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3662149"  
 /clone\_lib="NCI-CGAP\_Lu30"  
 /tissue\_type="tumor, metastatic to mammary"  
 /lab\_host="DH103"  
 /note="Organ: lung; Vector: PCMV-SpOrf6; Site\_1: NotI;  
 Site\_2: SalI; transgenic model WNT-1, expression driven by



MMTV-lTR enhancer; Cloned unidirectionally. Primer: Oligo  
 dr. Library constructed by Life Technologies.  
 Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT 130 a 60 c 88 g 140 t  
 ORIGIN

Query Match 4.2%; Score 20; DB 12; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 413 AATACCTGCCATTCCTC 432  
 |||||  
 Db 387 AATACCTGCCATTCCTC 368

RESULT 15  
 BG975401 420 bp mRNA linear EST 12-JUN-2001  
 LOCUS 602843052F1 NCI\_CGAP\_Man4 Mus musculus cDNA clone IMAGE:4978620 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG975401  
 VERSION BG975401.1 GI:14363038  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 420)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth  
 Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM10975 row: h column: 13  
 High quality sequence stop: 418.  
 Location/Qualifiers

FEATURES  
 source 1..420

/organism="Mus musculus"  
 /strain="NMRI"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4978620"  
 /clone\_lib="NCI\_CGAP\_Man4"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Priscilla Furth,  
 NIH Reference for transgenic model; Li et al., Cell Growth  
 and Differentiation 7, 3-11 (1996)."

BASE COUNT 149 a 87 c 58 g 125 t  
 ORIGIN 1 others

Query Match 4.2%; Score 20; DB 13; Length 420;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 413 AATACCTGCCATTCCTC 432  
 |||||  
 Db 22 AATACCTGCCATTCCTC 41

Search completed: July 20, 2003, 07:09:58  
 Job time: 1131 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:28:35 ; Search time 1535 Seconds

(without alignments)  
9005.754 Million cell updates/sec

Title: US-09-719-002-1

Perfect score: 475  
Sequence: 1 gaattctattgcgacctga.....acacacacacacacacatcaltg 475

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_lnu:\*  
20: em\_cm:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vit:\*  
38: em\_sy:\*  
39: em\_hgc\_hum:\*  
40: em\_hgc\_mus:\*  
41: em\_hgc\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	451	94.9	473	6 AX008564	AX008564 Sequence
2	53.4	11.2	240020	2 AC121587	AC121587 Mus muscu
3	53.4	11.2	289723	2 AC122935	AC122935 Mus muscu
4	52	10.9	171361	2 AC103286	AC103286 Rattus no
5	50.2	10.6	219342	10 AL732443	AL732443 Mouse DNA
6	49.6	10.4	16228	6 AX252126	AX252126 Sequence
7	49.6	10.4	16228	6 AX348891	AX348891 Sequence
8	48.6	10.2	37973	6 AX347098	AX347098 Sequence
9	48.4	10.2	15649	6 AX277940	AX277940 Sequence
10	48.4	10.2	15649	6 AX333627	AX333627 Sequence
11	47.8	10.1	34980	6 AX344550	AX344550 Sequence
12	47.4	10.0	8044	6 AX251380	AX251380 Sequence
13	46.4	9.8	549	3 DDAC1	X16525 Dictyostell
14	46.4	9.8	7532	6 AX345817	AX345817 Sequence
15	46.4	9.8	34980	6 AX344566	AX344566 Sequence
16	46.2	9.7	171146	6 AL161450	AL161450 Human DNA
17	46	9.7	5856	6 AX345471	AX345471 Sequence
18	46	9.7	6665	6 AX277841	AX277841 Sequence
19	46	9.7	6665	6 AX333516	AX333516 Sequence
20	46	9.7	6665	6 AX344985	AX344985 Sequence
21	46	9.7	101297	2 AC098899	AC098899 Rattus no
22	46	9.7	34980	6 AX344556	AX344556 Sequence
23	45.6	9.6	12592	6 AX251796	AX251796 Sequence
24	45.6	9.6	34980	6 AX344573	AX344573 Sequence
25	45.4	9.6	18987	6 AX345473	AX345473 Sequence
26	45.4	9.6	18987	6 AX347385	AX347385 Sequence
27	45.4	9.6	18987	6 AX349106	AX349106 Sequence
28	45.2	9.5	105783	6 AL162576	AL162576 Human DNA
29	45	9.5	6172	6 AX344634	AX344634 Sequence
30	45	9.5	8883	6 AX251519	AX251519 Sequence
31	45	9.5	48068	6 AL589883	AL589883 Human DNA
32	45	9.5	216724	5 AB073376	AB073376 Oryzias 1
33	44.8	9.4	5268	6 AX345779	AX345779 Sequence
34	44.8	9.4	17738	6 AX346440	AX346440 Sequence
35	44.8	9.4	321003	2 PFMALP3	AL035476 Plasmodiu
36	44.6	9.4	7352	6 AX345272	AX345272 Sequence
37	44.6	9.4	139990	2 AC095599	AC095599 Rattus no
38	44.6	9.4	272410	2 AC107492	AC107492 Rattus no
39	44.6	9.4	34980	6 AX344565	AX344565 Sequence
40	44.4	9.3	1141	6 AX083744	AX083744 Sequence
41	44.4	9.3	34980	6 AX344557	AX344557 Sequence
42	44.2	9.3	5768	6 AX347064	AX347064 Sequence
43	44	9.3	49981	6 AX059550	AX059550 Sequence
44	44	9.3	63292	8 F14G16	AF147260 Arabidops
45	44	9.3	152619	2 AC101920	AC101920 Mus muscu

## ALIGNMENTS

RESULT 1	AX008564	473 bp	DNA	linear	PAT 06-SEP-2000
LOCUS	AX008564				
DEFINITION	Sequence 1 from Patent WO9966057.				
ACCESSION	AX008564				
VERSION	AX008564.1	GI:9996114			
KEYWORDS					
SOURCE	garden asparagus.				
ORGANISM	Asparagus officinalis				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
	Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;				
	Asparagaceae; Asparagus.				
REFERENCE	1 (bases 1 to 473)				
AUTHORS	Draper, J., Kenton, P. and Paul, W.				
TITLE	Inducible promoters				

JOURNAL Patent: WO 9966057-A 1 23-DEC-1999; BIOGENMA UK LTD (GB); PAUL  
 DRAPER JOHN (GB); KENTON PAUL (GB);  
 WYATT (GB)

FEATURES  
 source 1. .473  
 Location/Qualifiers  
 /organism="Asparagus officinalis"  
 /db\_xref="taxon:4686"

BASE COUNT 162 a 112 c 67 g 132 t

ORIGIN

Query Match 94.9%; Score 451; DB 6; Length 473;  
 Best Local Similarity 99.6%; Pred. No. 4e-93;  
 Matches 473; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 GAATCTTATTCGACCTGACCTCTTGTGCTGCCGAGTGCTGCGAAATTCCTGT 60  
 |||||||  
 Db 1 GAATCTTATTCGACCTGACCTCTTGTGCTGCCGAGTGCTGCGAAATTCCTGT 60

QY 61 TCGCGACAACATAGTGTCTGCTGCTGATTTGACAGTTCACATATTTTCCATGTCAT 120  
 |||||||  
 Db 61 TCGCGACAACATAGTGTCTGCTGCTGATTTGACAGTTCACATATTTTCCATGTCAT 120

QY 121 GAGAGAGACATGACTAAATTAATTAATTCCTTAATCCCTAAACTCAATACAGAGAT 180  
 |||||||  
 Db 121 GAGAGAGACATGACTAAATTAATTAATTCCTTAATCCCTAAACTCAATACAGAGAT 180

QY 181 GACACATCGACAGAAAAATTTCTAATAGTCTTTCGCTGCTGTAAGATTTGAGAAAGATATAC 240  
 |||||||  
 Db 181 GACACATCGACAGAAAAATTTCTAATAGTCTTTCGCTGCTGTAAGATTTGAGAAAGATATAC 240

QY 241 CTACATTAATTAACAATTTTGCATTAATAATTAAGAACTTCAATACATGACATGCT 300  
 |||||||  
 Db 241 CTACATTAATTAACAATTTTGCATTAATAATTAAGAACTTCAATACATGACATGCT 300

QY 301 TCTAACATGAGACATGCTGACAGAACTGCTATTCCTTAATCCCAAGGCTTGACTTTCCA 360  
 |||||||  
 Db 301 TCTAACATGAGACATGCTGACAGAACTGCTATTCCTTAATCCCAAGGCTTGACTTTCCA 360

QY 361 CAATTCGAGATTAATCCATGAGCTGAGTACACATCCAAATTAATCCCTATAATACCTG 420  
 |||||||  
 Db 361 CAATTCGAGATTAATCCATGAGCTGAGTACACATCCAAATTAATCCCTATAATACCTG 420

QY 421 CCCATTCCCTCCCTCCAGACTCATCTAATCAAAAACACACACACACATCATG 475  
 |||||||  
 Db 421 CCCATTCCCTCCCTCCAGACTCATCTAATCAAAAACACACACACACATCATG 475

RESULT 2 240020 bp DNA linear HTG 25-MAY-2002  
 AC121587  
 LOCUS  
 DEFINITION  
 AC121587.11 unorderd pieces.  
 AC121587.2 GI:21206457  
 VERSION  
 AC121587.2 GI:21206457  
 KEYWORDS  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE  
 Mus musculus.  
 ORGANISM  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.  
 1 (bases 1 to 240020)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE  
 The sequence of Mus musculus clone  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 240020)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (20-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 240020)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA

## COMMENT

On May 25, 2002 this sequence version replaced gi:20986665.

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WTGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.wustl.edu  
 Project Information  
 Center project name: M\_BA0273002

## ----- Summary Statistics -----

Sequencing vector: M13; 0%  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-Primer ET; 0% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 234238 bases at least Q40  
 Consensus quality: 235156 bases at least Q30  
 Consensus quality: 235749 bases at least Q20  
 Insert size: 173000; agarose-fp  
 Insert size: 238837; sum-of-contigs  
 Quality coverage: 15.75 in Q20 bases; sum-of-contigs  
 Quality coverage: 9.77 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1565: contig of 1565 bp in length  
 1566 1565: gap of unknown length  
 1666 8142: contig of 6477 bp in length  
 8143 8242: gap of unknown length  
 8243 19060: contig of 10818 bp in length  
 19061 19160: gap of unknown length  
 19161 40614: contig of 21454 bp in length  
 40615 40715: gap of unknown length  
 40716 74844: contig of 34130 bp in length  
 74845 74945: gap of unknown length  
 74946 95947: contig of 21003 bp in length  
 95948 96047: gap of unknown length  
 96048 127209: contig of 31162 bp in length  
 127210 127310: gap of unknown length  
 127311 155476: contig of 28167 bp in length  
 155477 155576: gap of unknown length  
 155577 238360: contig of 82784 bp in length  
 238361 239059: gap of unknown length  
 239060 239159: contig of 609 bp in length  
 239160 240020: gap of unknown length  
 240021 240020: contig of 851 bp in length.

## FEATURES

## source

1. .240020  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP23-27302"  
 1. .1565  
 /note="assembly\_name:Contig56"  
 1666. .8142  
 /note="assembly\_name:Contig57"  
 8243. .19060  
 /note="assembly\_name:Contig58"  
 19161. .40614  
 /note="assembly\_name:Contig59"  
 40715. .74844  
 /note="assembly\_name:Contig60"  
 74945. .95947  
 /note="assembly\_name:Contig61"  
 96048. .127209  
 /note="assembly\_name:Contig62"

```

misc_feature      127310..155476
                  /note="assembly_name:Contig63"
misc_feature      155577..238360
                  /note="assembly_name:Contig64"
misc_feature      238461..239069
                  /note="assembly_name:Contig33"
misc_feature      239170..240020
                  /note="assembly_name:Contig25"
BASE COUNT      71540 a 49697 c 48772 g 68844 t 1167 others
ORIGIN
Query Match      11.2%; Score 53.4; DB 2; Length 240020;
Best Local Similarity 54.3%; Pred. No. 0.013;
Matches 108; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

OY      167  AATCAACAGCAGGATGACACATCCACACAAAAAATTCATATTAGTCCTTGGCTGTAGAAAT 226
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      67358  AAAAGAGACAAATGACATATATCTGAAAATTTATTTCCATAAATGTGTGTATATTT 67411
OY      227  TGGAACTGTAATACCTACATTAATTTACAACTTTTGCATAATAAATAAAGAAAGTTCTA 286
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      67418  TGTACAGAGGATATTTATTTAAATATTTTCAATGACATATTTTAAACAATATTTGA 67477
OY      287  ACAAGAGACTAGTCTTACACATGAGAGCTGTCCACGACTGCTACCTATTTCACCAAG 346
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      67478  ATAATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 67533
OY      347  GCTTAGACTTTCACCAAT 365
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      67538  GCGTAGTCACTGCGAAT 67556

RESULT 3
AC122935      289723 bp      DNA      linear      HTG 29-MAY-2002
LOCUS      Mus musculus chromosome UNK clone RP23-20B4, WORKING DRAFT
DEFINITION      SEQUENCE, 10 unordered pieces.
ACCESSION      AC122935
VERSION      AC122935.1 GI:21218568
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
ORGANISM      Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2. (bases 1 to 289723)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3. (bases 1 to 289723)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (29-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: M_BA0020E04
-----
Summary Statistics
-----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primed ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; 0.990319

```

```

Consensus quality: 285694 bases at least Q40
Consensus quality: 286828 bases at least Q30
Consensus quality: 287381 bases at least Q20
Insert size: 184000; agarose-ftp
Insert size: 294517; sum-of-contigs
Quality coverage: 21.66 in Q20 bases; agarose-ftp
Quality coverage: 9.89 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3055: contig of 3055 bp in length
*
* 3056: gap of unknown length
*
* 3156 7772: contig of 4617 bp in length
*
* 7773 7872: gap of unknown length
*
* 7873 11259: contig of 3387 bp in length
*
* 11260 11359: gap of unknown length
*
* 11360 16743: contig of 5384 bp in length
*
* 16744 16843: gap of unknown length
*
* 16844 25333: contig of 8490 bp in length
*
* 25334 25433: gap of unknown length
*
* 25434 37441: contig of 12008 bp in length
*
* 37442 37542: gap of unknown length
*
* 37542 73397: contig of 35856 bp in length
*
* 73398 73497: gap of unknown length
*
* 73498 118271: contig of 44774 bp in length
*
* 118272 118371: gap of unknown length
*
* 118372 217099: contig of 98788 bp in length
*
* 217100 217199: gap of unknown length
*
* 217200 289723: contig of 72524 bp in length.

FEATURES
Source
Location/Qualifiers
1..289723
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /chromosome="XUNK"
  /clone="RP23-20E4"
1..3055
  /note="assembly_name:Contig19"
  /note="assembly_name:Contig20"
  /note="assembly_name:Contig21"
  /note="assembly_name:Contig22"
  /note="assembly_name:Contig23"
  /note="assembly_name:Contig24"
  /note="assembly_name:Contig25"
  /note="assembly_name:Contig26"
  /note="assembly_name:Contig27"
  /note="assembly_name:Contig28"
  /note="assembly_name:Contig29"
  /note="assembly_name:Contig30"
  /note="assembly_name:Contig31"
  /note="assembly_name:Contig32"
  /note="assembly_name:Contig33"
  /note="assembly_name:Contig34"
  /note="assembly_name:Contig35"
  /note="assembly_name:Contig36"
  /note="assembly_name:Contig37"
  /note="assembly_name:Contig38"
  /note="assembly_name:Contig39"
  /note="assembly_name:Contig40"
  /note="assembly_name:Contig41"
  /note="assembly_name:Contig42"
  /note="assembly_name:Contig43"
  /note="assembly_name:Contig44"
  /note="assembly_name:Contig45"
  /note="assembly_name:Contig46"
  /note="assembly_name:Contig47"
  /note="assembly_name:Contig48"
  /note="assembly_name:Contig49"
  /note="assembly_name:Contig50"
  /note="assembly_name:Contig51"
  /note="assembly_name:Contig52"
  /note="assembly_name:Contig53"
  /note="assembly_name:Contig54"
  /note="assembly_name:Contig55"
  /note="assembly_name:Contig56"
  /note="assembly_name:Contig57"
  /note="assembly_name:Contig58"
  /note="assembly_name:Contig59"
  /note="assembly_name:Contig60"
  /note="assembly_name:Contig61"
  /note="assembly_name:Contig62"
  /note="assembly_name:Contig63"
  /note="assembly_name:Contig64"
  /note="assembly_name:Contig65"
  /note="assembly_name:Contig66"
  /note="assembly_name:Contig67"
  /note="assembly_name:Contig68"
  /note="assembly_name:Contig69"
  /note="assembly_name:Contig70"
  /note="assembly_name:Contig71"
  /note="assembly_name:Contig72"
  /note="assembly_name:Contig73"
  /note="assembly_name:Contig74"
  /note="assembly_name:Contig75"
  /note="assembly_name:Contig76"
  /note="assembly_name:Contig77"
  /note="assembly_name:Contig78"
  /note="assembly_name:Contig79"
  /note="assembly_name:Contig80"
  /note="assembly_name:Contig81"
  /note="assembly_name:Contig82"
  /note="assembly_name:Contig83"
  /note="assembly_name:Contig84"
  /note="assembly_name:Contig85"
  /note="assembly_name:Contig86"
  /note="assembly_name:Contig87"
  /note="assembly_name:Contig88"
  /note="assembly_name:Contig89"
  /note="assembly_name:Contig90"
  /note="assembly_name:Contig91"
  /note="assembly_name:Contig92"
  /note="assembly_name:Contig93"
  /note="assembly_name:Contig94"
  /note="assembly_name:Contig95"
  /note="assembly_name:Contig96"
  /note="assembly_name:Contig97"
  /note="assembly_name:Contig98"
  /note="assembly_name:Contig99"
  /note="assembly_name:Contig100"
  /note="assembly_name:Contig101"
  /note="assembly_name:Contig102"
  /note="assembly_name:Contig103"
  /note="assembly_name:Contig104"
  /note="assembly_name:Contig105"
  /note="assembly_name:Contig106"
  /note="assembly_name:Contig107"
  /note="assembly_name:Contig108"
  /note="assembly_name:Contig109"
  /note="assembly_name:Contig110"
  /note="assembly_name:Contig111"
  /note="assembly_name:Contig112"
  /note="assembly_name:Contig113"
  /note="assembly_name:Contig114"
  /note="assembly_name:Contig115"
  /note="assembly_name:Contig116"
  /note="assembly_name:Contig117"
  /note="assembly_name:Contig118"
  /note="assembly_name:Contig119"
  /note="assembly_name:Contig120"
  /note="assembly_name:Contig121"
  /note="assembly_name:Contig122"
  /note="assembly_name:Contig123"
  /note="assembly_name:Contig124"
  /note="assembly_name:Contig125"
  /note="assembly_name:Contig126"
  /note="assembly_name:Contig127"
  /note="assembly_name:Contig128"
  /note="assembly_name:Contig129"
  /note="assembly_name:Contig130"
  /note="assembly_name:Contig131"
  /note="assembly_name:Contig132"
  /note="assembly_name:Contig133"
  /note="assembly_name:Contig134"
  /note="assembly_name:Contig135"
  /note="assembly_name:Contig136"
  /note="assembly_name:Contig137"
  /note="assembly_name:Contig138"
  /note="assembly_name:Contig139"
  /note="assembly_name:Contig140"
  /note="assembly_name:Contig141"
  /note="assembly_name:Contig142"
  /note="assembly_name:Contig143"
  /note="assembly_name:Contig144"
  /note="assembly_name:Contig145"
  /note="assembly_name:Contig146"
  /note="assembly_name:Contig147"
  /note="assembly_name:Contig148"
  /note="assembly_name:Contig149"
  /note="assembly_name:Contig150"
  /note="assembly_name:Contig151"
  /note="assembly_name:Contig152"
  /note="assembly_name:Contig153"
  /note="assembly_name:Contig154"
  /note="assembly_name:Contig155"
  /note="assembly_name:Contig156"
  /note="assembly_name:Contig157"
  /note="assembly_name:Contig158"
  /note="assembly_name:Contig159"
  /note="assembly_name:Contig160"
  /note="assembly_name:Contig161"
  /note="assembly_name:Contig162"
  /note="assembly_name:Contig163"
  /note="assembly_name:Contig164"
  /note="assembly_name:Contig165"
  /note="assembly_name:Contig166"
  /note="assembly_name:Contig167"
  /note="assembly_name:Contig168"
  /note="assembly_name:Contig169"
  /note="assembly_name:Contig170"
  /note="assembly_name:Contig171"
  /note="assembly_name:Contig172"
  /note="assembly_name:Contig173"
  /note="assembly_name:Contig174"
  /note="assembly_name:Contig175"
  /note="assembly_name:Contig176"
  /note="assembly_name:Contig177"
  /note="assembly_name:Contig178"
  /note="assembly_name:Contig179"
  /note="assembly_name:Contig180"
  /note="assembly_name:Contig181"
  /note="assembly_name:Contig182"
  /note="assembly_name:Contig183"
  /note="assembly_name:Contig184"
  /note="assembly_name:Contig185"
  /note="assembly_name:Contig186"
  /note="assembly_name:Contig187"
  /note="assembly_name:Contig188"
  /note="assembly_name:Contig189"
  /note="assembly_name:Contig190"
  /note="assembly_name:Contig191"
  /note="assembly_name:Contig192"
  /note="assembly_name:Contig193"
  /note="assembly_name:Contig194"
  /note="assembly_name:Contig195"
  /note="assembly_name:Contig196"
  /note="assembly_name:Contig197"
  /note="assembly_name:Contig198"
  /note="assembly_name:Contig199"
  /note="assembly_name:Contig200"
  /note="assembly_name:Contig201"
  /note="assembly_name:Contig202"
  /note="assembly_name:Contig203"
  /note="assembly_name:Contig204"
  /note="assembly_name:Contig205"
  /note="assembly_name:Contig206"
  /note="assembly_name:Contig207"
  /note="assembly_name:Contig208"
  /note="assembly_name:Contig209"
  /note="assembly_name:Contig210"
  /note="assembly_name:Contig211"
  /note="assembly_name:Contig212"
  /note="assembly_name:Contig213"
  /note="assembly_name:Contig214"
  /note="assembly_name:Contig215"
  /note="assembly_name:Contig216"
  /note="assembly_name:Contig217"
  /note="assembly_name:Contig218"
  /note="assembly_name:Contig219"
  /note="assembly_name:Contig220"
  /note="assembly_name:Contig221"
  /note="assembly_name:Contig222"
  /note="assembly_name:Contig223"
  /note="assembly_name:Contig224"
  /note="assembly_name:Contig225"
  /note="assembly_name:Contig226"
  /note="assembly_name:Contig227"
  /note="assembly_name:Contig228"
  /note="assembly_name:Contig229"
  /note="assembly_name:Contig230"
  /note="assembly_name:Contig231"
  /note="assembly_name:Contig232"
  /note="assembly_name:Contig233"
  /note="assembly_name:Contig234"
  /note="assembly_name:Contig235"
  /note="assembly_name:Contig236"
  /note="assembly_name:Contig237"
  /note="assembly_name:Contig238"
  /note="assembly_name:Contig239"
  /note="assembly_name:Contig240"
  /note="assembly_name:Contig241"
  /note="assembly_name:Contig242"
  /note="assembly_name:Contig243"
  /note="assembly_name:Contig244"
  /note="assembly_name:Contig245"
  /note="assembly_name:Contig246"
  /note="assembly_name:Contig247"
  /note="assembly_name:Contig248"
  /note="assembly_name:Contig249"
  /note="assembly_name:Contig250"
  /note="assembly_name:Contig251"
  /note="assembly_name:Contig252"
  /note="assembly_name:Contig253"
  /note="assembly_name:Contig254"
  /note="assembly_name:Contig255"
  /note="assembly_name:Contig256"
  /note="assembly_name:Contig257"
  /note="assembly_name:Contig258"
  /note="assembly_name:Contig259"
  /note="assembly_name:Contig260"
  /note="assembly_name:Contig261"
  /note
```



```

* 31170 31269: gap of unknown length
* 31270 33415: contig of 2146 bp in length
* 33416 33515: gap of unknown length
* 33516 35710: contig of 2195 bp in length
* 35711 35810: gap of unknown length
* 35811 37964: contig of 2154 bp in length
* 37965 38064: gap of unknown length
* 38065 40086: contig of 2022 bp in length
* 40087 40186: gap of unknown length
* 40187 41776: contig of 1590 bp in length
* 41777 41876: gap of unknown length
* 41877 44079: contig of 2203 bp in length
* 44080 44179: gap of unknown length
* 44180 47089: contig of 2910 bp in length
* 47090 47189: gap of unknown length
* 47190 49139: contig of 1950 bp in length
* 49140 49239: gap of unknown length
* 49240 51349: contig of 2010 bp in length
* 51350 51350: gap of unknown length
* 51350 54113: contig of 2764 bp in length
* 54114 54213: gap of unknown length
* 54214 56805: contig of 2592 bp in length
* 56806 56905: gap of unknown length
* 56906 59181: contig of 2276 bp in length
* 59182 59281: gap of unknown length
* 59282 61685: contig of 2404 bp in length
* 61686 61785: gap of unknown length
* 61786 65961: contig of 4176 bp in length
* 65962 66061: gap of unknown length
* 66062 69179: contig of 3118 bp in length
* 69180 69279: gap of unknown length
* 69280 71949: contig of 2670 bp in length
* 71950 72049: gap of unknown length
* 72050 74860: contig of 2811 bp in length
* 74861 74960: gap of unknown length
* 74961 77963: contig of 3003 bp in length
* 77964 78064: gap of unknown length
* 78065 80491: contig of 2428 bp in length
* 80492 80591: gap of unknown length
* 80592 84117: contig of 3526 bp in length
* 84118 84217: gap of unknown length
* 84218 88466: contig of 4429 bp in length
* 88467 88747: gap of unknown length
* 88747 91528: contig of 2762 bp in length
* 91529 91628: gap of unknown length
* 91629 95378: contig of 3750 bp in length
* 95379 95478: gap of unknown length
* 95479 99361: contig of 3883 bp in length
* 99362 99461: gap of unknown length
* 99462 103099: contig of 3638 bp in length
* 103100 103199: gap of unknown length
* 103200 107950: contig of 4731 bp in length
* 107951 108030: gap of unknown length
* 108031 111476: contig of 3446 bp in length
* 111477 111576: gap of unknown length
* 111577 115661: contig of 4085 bp in length
* 115662 115761: gap of unknown length
* 115762 119386: contig of 3623 bp in length
* 119387 119486: gap of unknown length
* 119487 124224: contig of 4738 bp in length
* 124225 124324: gap of unknown length
* 124325 128353: contig of 4029 bp in length
* 128354 128453: gap of unknown length
* 128454 133606: contig of 5153 bp in length
* 133607 133706: gap of unknown length
* 133707 138517: contig of 4811 bp in length
* 138518 142807: gap of unknown length
* 138618 142807: contig of 4190 bp in length

```

```

Query Match      10.9% Score 52; DB 2; Length 171361;
Best Local Similarity 52.3% Pred No. 0.029;
Matches 115; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```

```

QY 1 GAATTCATTGCGACCTGACTCTCTGTTGCTGCCGAGCTGCTCGAAATTCGT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133811 GGATTCGCTTTGTGGCAGCTGTGCTCTCTTATGCGACGATGCTGTTAGACGCGAGT 133870
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 TGGCGACACACTACTGGGCCCTTGCTGATTGGAAGTCCCAATATATTCATGTCAT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133871 TACTAGAGAGTGTGCTGTTGTTCTTACTTAAATATCACTGAACTTAAATAT 133930
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GAGAGAAAGCAGTACTGCTAAAGTAAATAGCTTAATCCCTAAACATCAATCAAGAT 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133931 GAGATATGAAATAGTTTGCAGAGTATCTAGACCGCATGCTAGCTAAGTAAATGATA 133990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GACACATCCACAGAAAAATTCATTAATTAAGTTTGGCTGT 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 133991 GACTCTGACACATCTTACAAATTTCAATCAATCAATCCATGCT 134030
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 5
AL732443/c
LOCUS AL732443 219342 bp DNA linear ROD 02-JUL-2002
DEFINITION Mouse DNA sequence from clone RP23-90H6 on chromosome X, complete
sequence.
ACCESSION AL732443
VERSION AL732443.6 GI:21531574
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 219342)
AUTHORS Kay,M.
TITLE Direct Submission
COMMENT Submitted (02-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
On Jun 21, 2002 this sequence version replaced gi:21314875.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquyer@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPep; Information on the WormPep database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-90H6 is from the RPCR-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

#### FEATURES

##### source

```

1. 219342
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-90H6"
/clone_lib="RPCR-23"
BASE COUNT 66074 a 38077 c 38575 g 76616 t
ORIGIN

```

[illegible]

```

DEFINITION      Sequence 349 from patent WO0202807.
ACCESSION       AX348891
VERSION         AX348891.1  GI:18614926
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS        Olak A., Piepenbrock C. and Berlin R.
TITLE           Diagnosis of diseases associated with cell signalling
JOURNAL         Patent: NO 0202807-A 349 10-JAN-2002;
                EpiGenomics AG (DE)
FEATURES
Source
    1..16228
    /organism="synthetic construct"
    /db_xref="taxon:32630"
    /note="Chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      4652 a      198 c      3601 g      7777 t
ORIGIN

```

RESULT 6	AX252126/c	AX252126	16228 bp	DNA	linear	PAT 05-OCT-2001
LOCUS						
DEFINITION		Sequence	387	from Patent WO0168911.		
ACCESSION		AX252126				
VERSION		AX252126.1		GI:15985481		
KEYWORDS						
SOURCE		synthetic construct.				
ORGANISM		synthetic construct				
		artificial sequences.				
REFERENCE		1 (bases 1 to 16228)				
AUTHORS		Olek, A., Piepenbrock, C. and Berlin, K.				
TITLE		Diagnosis of diseases associated with the cell cycle				
JOURNAL		Patent: WO 0168911-A 387 20-SEP-2001;				
		Epigenomics AG (DE)				
FEATURES		Location/Qualifiers				
source		1..16228				
		/organism="synthetic construct"				
		/db_xref="taxon:32630"				
		/note="Chemically treated genomic DNA (Homo sapiens)"				
BASE COUNT		4652 a	198 c	3601 g	7777 t	
ORIGIN						

Query Match	10.4%	Score 49.6	DB 6	Length 16228
Best Local Similarity	48.3%	Pred. Mismatches 0.15		
Matches 139	Conservative	0	Mismatches 149	Indels 0
			Gaps 0	
QY	131	CATGACTAAAGTAATTAGCTTAATCCCTAAAACTCAATACAAACGAGATGACATCCA	190	
Db	9241	CTTTCTATTATTTAAACCAAACTTAAATTTTAAAAAATAAACTAATATATATA	9182	
QY	191	CAGAAAAATTCCTATTASTCTTTCGCGTGAATAAATGTGAAACTAATACTACATTAT	250	
Db	9181	TAAAAAAATATATACCTACATTTTCGCACATAAACCTTCAAACTTCAATATTATTAT	9122	
QY	251	TACAACTTTGCCAATAAATAATAAGAAAGTTCTACATGAGACCTAGTTCTAACATGA	310	
Db	9121	ATACTATATACACAAATTAATATTATAATACCTCAAAATATATTATTTAAACTATT	9062	
QY	311	AACATAGTCACAGAACTCGTACCTTATTCACAAAGGCTTAGCTTCCACAAATCGGA	370	
Db	9061	AACCAACGCGATACCTACTACTCTCAATCCCAACACTTTAAAAAACCACAAATAAAC	9002	
QY	371	TTATGCCATGAGCTGATGACACCACTCCAAATTATCCCTATATAATACC	418	
Db	9001	TCACGAAATCAAAAATAAAAAACCATCTCAACCAACATATATAAAACC	8954	

Query Match	Score	DB	Length
Best Local Similarity 48.3%	Pred. No. 0.22	164	Indels 1; Gaps 1;
Matches 158; Conservative 0; Mismatches 164; Indels 1; Gaps 1;			

	Query Match	10.2%	Score 48.6	DB 6	Length 37973
	Best Local Similarity	48.9%	Pred No. 0.22		
	Matches 156	Conservative	0	Mismatches 164	Indels 1
07	130 ACATGACTAAGTAATTAGCTTATCCGCTAAACCTCATTACCAACGAGATGCACATCC 189				

Db 26628 AATAAATAAATAATCATCAAAACAAAAAATTTCAATATACACACTA 26569  
QY 190 ACAGAAAAAATTTCTAATTGCTTGGTGTAGAAATTTGGAATCGAATACCTACATTAA 249  
Db 26568 AAAACAAAAAATACAAAAAATTAATTAACAAAAATCCAAAAAATCCTACTATCA 26509  
QY 250 TTACAACTTTTGCATTAATAATATAAGAAAGTTCACATGAGACTAGTCTTAAGT 309  
Db 26508 CCGTAA-ACATCAAACTTAATAATAATTTATTTACTTAACATATTTCAAAAAATTC 26450  
QY 310 AAGACTAGTCCAGAACTGTACTTATTCACAAAGGCTTAGACTTTCCACAAATCGAG 369  
Db 26449 TAACCGAATACATTAATCTGTACCTAATATCCCAACACTTTAAAAACCGAAACGTAA 26390  
QY 370 ATTATCCATGAGACTGATGACACCTCAATTAATCCCTAATTAATACCTGCCATGCC 429  
Db 26389 ATCAGAAATCAAAAAATGGAACCACTCTTAACACACATTAATAACCCCGTCTCTCTA 26330  
QY 430 CTCCTCCAGACTCATCTAATCA 452  
Db 26329 AAAATACAAAAATTAATTAACGA 26307

RESULT 9  
AX277940/c 15649 bp DNA linear PAT 01-NOV-2001  
LOCUS AX277940 Sequence 103 from Patent WO0177375.  
DEFINITION AX277940  
ACCESSION AX277940  
VERSION AX277940.1 GI:16605012  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequences.  
REFERENCE 1  
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with gene regulation  
JOURNAL Patent: WO 0177375-A 103 18-OCT-2001;  
Epigenomics AG (DE)  
FEATURES  
source location/Qualifiers  
1..15649  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 4082 a 261 c 3731 g 7575 t  
ORIGIN

Query Match 10.2%; Score 48.4; DB 6; Length 15649;  
Best Local Similarity 46.7%; Pred. No. 0.28;  
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 136 CTAAGTAAATAGCTTAATCCCTTAACCTCAATGCAAGAGATGACATCCACAGAA 195  
Db 11798 CTAACACAAACCACTACTCTTCCACACTCCCAAAAAATCAATTTTTCATATAAAAAACA 11739  
QY 196 AAAATTTCTAATAGCTTTCGCTGAGAAATTTGGAATCGAATCAATTAATTAACA 255  
Db 11738 ACAAACCTTTAAACACATACTATCTTCTTAATAATTAATAAAACATAAAAAATA 11679  
QY 256 CTTTGGCAATTAATAATAAGAAAGTCTTAACATGAGAGATGCTTAACATGAAGCT 315  
Db 11678 CATTAACTTTCAATTTTAACACATTTAAACACATCTTAACCTCAAACTCAAAAT 11619  
QY 316 AGTCCAGAACTGCTACTTAATTCACAAAGGCTTAGACTTCCACAAATGAGATTATC 375  
Db 11618 TTTTCTTTCATTTTCTTTTAAACAAATTTGCTCTATTCCTCCCAACTAAATAACA 11559  
QY 376 CCATGAGCTGATGACACCAATATATCCATTAATATCTGCCCATCCCTCTCTC 435  
Db 11558 ATACGATATCTCAACTACATCACTCCGCTCTTAATTAATCAAACTTCTCTACTT 11499  
QY 436 CAGACTCATCTAATCAAAACACACACA 465  
Db 11498 CAACCTCTTAATTAATCAAACTCAATAA 11469

Db 11498 CAACCTCTTAATTAATCAAACTCAATAA 11469

RESULT 10  
AX323627/c 15649 bp DNA linear PAT 07-JAN-2002  
LOCUS AX323627 Sequence 115 from Patent WO0192565.  
DEFINITION AX323627  
ACCESSION AX323627  
VERSION AX323627.1 GI:18094375  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequences.  
REFERENCE 1  
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of diseases associated with dna transcription  
JOURNAL Patent: WO 0192565-A 115 06-DEC-2001;  
Epigenomics AG (DE)  
FEATURES  
source location/Qualifiers  
1..15649  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 4082 a 261 c 3731 g 7575 t  
ORIGIN

Query Match 10.2%; Score 48.4; DB 6; Length 15649;  
Best Local Similarity 46.7%; Pred. No. 0.28;  
Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 136 CTAAGTAAATAGCTTAATCCCTTAACCTCAATGCAAGAGATGACATCCACAGAA 195  
Db 11798 CTAACACAAACCACTACTCTTCCACACTCCCAAAAAATCAATTTTTCATATAAAAAACA 11739  
QY 196 AAAATTTCTAATAGCTTTCGCTGAGAAATTTGGAATCGAATCAATTAATTAACA 255  
Db 11738 ACAAACCTTTAAACACATACTATCTTAAATAATTAATAAAACATAAAAAATA 11679  
QY 256 CTTTGGCAATTAATAATAAGAAAGTCTTAACATGAGAGCTTCTTAACATGAAGACT 315  
Db 11678 CATTAACTTTCTAATTTTAACACATTAATTAACACATCTCAACCTCAAACTCAAAAT 11619  
QY 316 AGTCCAGAACTGCTACTTAATTCACAAAGGCTTAGACTTCCACAAATGAGATTATC 375  
Db 11618 TTTTCTTTCATTTTCTTTTAAACAAATTTGCTCTATTCCTCCCAACTCAAAAT 11559  
QY 376 CCATGAGCTGATGACACCAATTAATTCCTTAATTAATCTGCCCATCCCTCTCTC 435  
Db 11558 ATACGATATCTCAACTACATCACTCCGCTCTTAATTAATCAAACTTCTCTACTT 11499  
QY 436 CAGACTCATCTAATCAAAACACACACA 465  
Db 11498 CAACCTCTTAATTAATCAAACTCAATAA 11469

RESULT 11  
AX344550/c 34980 bp DNA linear PAT 01-FEB-2002  
LOCUS AX344550 Sequence 1 from Patent WO0200932.  
DEFINITION AX344550  
ACCESSION AX344550  
VERSION AX344550.1 GI:18492436  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
artificial sequences.  
REFERENCE 1  
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.  
TITLE Diagnosis of known genetic parameters within the mhc  
JOURNAL Patent: WO 0200932-A 1 03-JAN-2002;  
Epigenomics AG (DE)  
FEATURES  
source location/Qualifiers  
1..34980  
/organism="synthetic construct"



```

/db.xref="taxon:32630"
/note="Chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3,673,778 <223>-split as follows:-seq 01 0.000.001 to 0.349,980-seq 02 0.300.001 649,980-seq 03 600.001 949,980-seq 04 900.001 1,249,980-seq 05 1,200.001 1,549,980-seq 06 1,500.001 1,849,980-seq 07 1,800.001 2,149,980-seq 08 2,100.001 2,449,980-seq 09 2,400.001 2,749,980-seq 10 2,700.001 3,049,980-seq 11 3,000.001 3,349,980-seq 12 3,300.001 3,649,980-seq 13 3,600.001 3,673,778 <223>-Original length of seq 2: 3,673,778 <223>-split as follows:-seq 14 0.000.001 to 0.349,980-seq 15 0.300.001 649,980-seq 16 600.001 949,980-seq 17 900.001 1,249,980-seq 18 1,200.001 1,549,980-seq 19 1,500.001 1,849,980-seq 20 1,800.001 2,149,980-seq 21 2,100.001 2,449,980-seq 22 2,400.001 2,749,980-seq 23 2,700.001 3,049,980-seq 24 3,000.001 3,349,980-seq 25 3,300.001 3,649,980-seq 26 3,600.001 3,673,778"

BASE COUNT      100428 a      4240 c      79243 g      166069 t
ORIGIN

Query Match
Best Local Similarity 10.1%; Score 47.8; DB 6; Length 349980;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 160 AAACTCAATGAAACGAGATGACATCCACAGAAAATTGTAATTGCTTGGGTG 219
Pb 167784 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167725
QY 220 TAGAATATGGAACCTGAACTGATTAATTAACACTTTGCAATTAATTAAGAA 279
Pb 167724 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 167665
QY 280 AGTTCTACATGAGACTAGTTCTTAACATGAGACTGTCCACAGACTGCTGACTTTC 339
Pb 167664 AATTTAAAAAAGAAAAAAGAAAAAAGAAAAAATTAATTAATTAATTAATTAAT 167605
QY 340 CACAAAGGCTTAGACTTTCACAAATGAGATTAATCCAGACTGATGACACCAATCA 399
Pb 167604 CCAACACTTTAAAAAATTAACAAACAAATCAATTAATTAATTAATTAATTAAT 167545
QY 400 AATATCCCTTAATATACCTGCCCATGCCCTCCCTCCAGACTGCTTAATCAAAAACA 459
Pb 167544 AACTTACATATTAATTAACCCCTCTCTACTAAAAATTAACAAAAATTAACATAATA 167485
QY 460 CACACACCAATCAAT 474
Pb 167484 ACAATACCTTAATAT 167470

RESULT 12
AX251380/c 8044 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 348 from Patent WO0168912.
ACCESSION AX251380
VERSION AX251380.1 GI:15984803
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 8044)
AUTHORS Olek,A., Pilepdrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncosenes
JOURNAL Patent: WO 0168912-A 348 20-SEP-2001;
Epidemiology AG (DE)
FEATURES
SOURCE location/Qualifiers
1..8044
/organism="synthetic construct"
/db.xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      2017 a      211 c      1923 g      3893 t
ORIGIN

```

```

Query Match
Best Local Similarity 10.0%; Score 47.4; DB 6; Length 8044;
Matches 168; Conservative 0; Mismatches 181; Indels 2; Gaps 1;

QY 124 AGAAGCAGATGATTAAGTATTAGTTAATCCCTAAAACCTCAATACAGAGATGAC 183
Pb 1413 AAAATCATAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1354
QY 184 ACATCCACAGAAAAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 243
Pb 1353 ATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1294
QY 244 CATTAAATTAACACTTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 303
Pb 1293 CAATTAATTAACCAACTATCTACCTAAATTAATTAATTAATTAATTAATTAATTAAT 1234
QY 304 AACATGAGACTGTCCAGCACTGCTTACTTATTCGAAAGAGCTTACACTTTCCACA 363
Pb 1233 AACTTATTAACGAAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1174
QY 364 ATGAGATTATCCCATGAGACTGA--TGACACCATCCAAATTAATCCCTAATTAATCCGC 421
Pb 1173 AACCAATTAACCTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1114
QY 422 CCATTCCTCCCTCCGAGACTCATCTTACTCAAAAACACACACACCAATTC 472
Pb 1113 CTCTACTAAATAATCAAAATTAACCAAAATTAATTAATTAATTAATTAATTAATTAAT 1063

RESULT 13
DDAACL 549 bp mRNA linear INV 05-AUG-1995
DEFINITION Dictyostellium discoideum AAC-rich mRNA (AACL).
ACCESSION X16525
VERSION X16525.1 GI:7172
KEYWORDS threonine rich protein.
SOURCE Dictyostellium discoideum.
ORGANISM Dictyostellium discoideum.
REFERENCE 1 (bases 1 to 549)
AUTHORS Shaw,D.R., Richter,H., Giorda,R., Omachi,T. and Emis,H.L.
TITLE Nucleotide sequences of Dictyostellium discoideum developmentally regulated cDNAs rich in (AAC) imply proteins that contain clusters of asparagine, glutamine, or threonine
JOURNAL Mol. Gen. Genet. 218 (3), 453-459 (1989)
MEDLINE 90066348
PUBMED 2511421
FEATURES
SOURCE location/Qualifiers
1..549
/organism="Dictyostellium discoideum"
/db.xref="taxon:44689"
/clone_lib="AAC1"
<1..>549
/note="coding region (AA 1 - 183)"
/odon_start=1
/protein_id="CAA34532.1"
/db.xref="GI:930011"
/db.xref="SWISS-PROT:P14195"
/translacion="DNDNNSSTRPKQCKLKSNESTSTSTTTTPPIITTTTTTTTT
TPNLSKYNYLNLYIEKONKMLPTTEETETITPTTTTITTTTITTTTITTTTITTTTITTT
HPLPIQKEITFLVELVGSFPMARKVKYKKKVCNGEVENLNIYFDIHLSAVKHV
SEPVKLSNSDFHQSVPFNG"
misc.feature 67..135
/note="threonine cluster"
misc.feature 204..270
/note="threonine cluster"
BASE COUNT      264 a      98 c      45 g      142 t
ORIGIN
Query Match
Best Local Similarity 9.8%; Score 46.4; DB 3; Length 549;
Matches 148; Conservative 0; Mismatches 181; Indels 2; Gaps 1;

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:02:54 ; Search time 191 seconds  
(without alignments)  
5600.519 Million cell updates/sec

Title: US-09-719-002-1

Perfect score: 475

Sequence: 1 gaattctattgcagctga.....aacacacacacacatcatg 475

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N\_Geneseq.101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	100.0	475	AAZ29510	A.officinalis than
2	49.6	10.4	16228	ABL70459	Chemically treated
3	49.6	10.4	16228	AA561424	Human gene regulat
4	48.6	10.2	37873	ABL31196	Human immune syste
5	48.4	10.2	15649	AA545396	Chemically pretrea
6	48.4	10.2	15649	ABK28241	DNA transcription
7	47.4	10.0	8044	AA546626	Tumour suppressor
8	46.4	9.8	7522	ABL32915	Human immune syste
9	46	9.7	5856	ABL32559	Human immune syste

C 10	46	9.7	6665	22	AA545299	Chemically pretrea
C 11	46	9.7	6665	24	ABL22083	Human immune syste
C 12	46	9.7	6665	24	ABK28130	DNA transcription
C 13	45.6	9.6	12592	24	AA561101	Human gene regulat
C 14	45.4	9.6	18997	24	ABK33949	Human DNA for stag
C 15	45.4	9.6	18997	24	ABL32571	Human immune syste
C 16	45	9.5	6172	24	ABN80042	Human chemically m
C 17	45	9.5	8883	22	AA546761	Tumour suppressor
C 18	44.8	9.4	5268	24	ABL32897	Human immune syste
C 19	44.8	9.4	17738	24	ABL33538	Human immune syste
C 20	44.6	9.4	7352	24	ABL32370	Human immune syste
C 21	44.2	9.3	5768	24	ABL34162	Human immune syste
C 22	44	9.3	163319	21	AA522306	Arabidopsis thalia
C 23	43.8	9.2	9646	24	ABL33688	Human immune syste
C 24	43.6	9.2	6162	24	ABL92315	Chemically treated
C 25	43.6	9.2	14253	24	ABL33495	Human immune syste
C 26	43.4	9.1	5557	24	ABL33547	Human immune syste
C 27	43.4	9.1	8622	24	ABL34183	Human immune syste
C 28	43.4	9.1	11944	24	ABL34187	Human immune syste
C 29	43.2	9.1	11812	22	AA545502	Chemically pretrea
C 30	43.2	9.1	11812	22	AA546742	Tumour suppressor
C 31	43.2	9.1	11812	22	ABL34119	Human immune syste
C 32	43.2	9.1	11812	24	ABK28432	DNA transcription
C 33	42.8	9.0	6694	24	ABL33685	Human immune syste
C 34	42.8	9.0	513445	22	AA161373	Soybean 318013 reg
C 35	42.6	9.0	3920	24	ABO67030	Human anglogenesis
C 36	42.6	9.0	6290	24	ABL33047	Human immune syste
C 37	42.6	9.0	10717	24	ABN80210	Human immune syste
C 38	42.6	9.0	10717	24	ABL33694	Human chemically m
C 39	42.6	9.0	33053	24	ABO67006	Human immune syste
C 40	42.4	8.9	5928	24	ABL33059	Human anglogenesis
C 41	42.4	8.9	6866	24	ABL49320	Human immune syste
C 42	42.4	8.9	6866	24	ABL32667	Human polynucleoti
C 43	42.4	8.9	15732	22	AA545389	Human immune syste
C 44	42.4	8.9	15732	22	AA545389	Chemically pretrea
C 45	42.4	8.9	83391	24	ABK28234	DNA transcription
					ABO67094	Human anglogenesis

## ALIGNMENTS

RESULT 1	AAZ29510	standard; DNA: 475 BP.
ID	AAZ29510	
AC	AAZ29510;	
XX	14-MAR-2000	(first entry)
XX	A.officinalis thaumatin-like PR-5 related gene (Aoprt-L) promoter.	
DT	Inducible promoter; Thaumatin-like PR-5 related gene; Aoprt-L; trait;	
XX	non-phytoxic inducing agent; Salicylic acid; SA; BTH; transformed plant;	
XX	systemic activation; developmental; environmental; pathogen resistance;	
KW	heterologous gene; disease control; sterility; fertility;	
KW	fruit ripening; ds.	
XX	Asparagus officinalis.	
XX	OS	
XX	Key	
XX	misc_feature	
XX	100..111	
XX	Location/Qualifiers	
XX	/tag- a	
XX	/note- "Homologous to Potato wound induced promoter	
XX	sequence"	
XX	205..220	
XX	misc_feature	
XX	/tag- b	
XX	/note- "Homologous to Tobacco PR-2 promoter sequence"	
XX	226..341	
XX	misc_signal	
XX	/tag- c	
XX	/label- "SA responsive element	
XX	/note- "Induced by non-phytoxic agents like Salicylic	
XX	acid or BTH"	
XX	251..264	
XX	misc_feature	

FT /\*tag= d  
 FT /note= "Homologous to Carrot PR-3 and PR-4 promoter  
 FT sequence"  
 FT repeat\_region 281..318  
 FT /tag= e  
 FT /rpt\_type= TANDEM  
 FT repeat\_unit 281..298  
 FT /tag= f  
 FT /note= "18 bp repeat"  
 FT 409..416  
 FT /tag= g

W09966057-A2.

23-DEC-1999.

21-JUN-1999; 99WO-GB01949.

19-JUN-1998; 98GB-0013345.

(BIOG-) BIOEMMA UK LTD.

Draper J, Kenton P, Paul W;

WPI; 2000-106107/09.

Novel promoters used to control the expression of heterologous genes in transformed plants

Claim 4; Fig 6; 67pp; English.

The present DNA sequence is a novel inducible promoter, derived from Asparagus officinalis thumatin-like PR-5 related gene, AopPT-L. The promoter sequence has similarities with other PR promoters. The promoter is responsive to low levels of an environmentally-acceptable and non-phytoxic inducing agents, like salicylic acid or BTH. The promoters also exhibit low levels of pathogen induced systemic activation and environmentally or developmentally induced expression. The inducible promoter sequence is used to control the expression of heterologous genes in transformed plants, especially genes whose products affect a trait of the plant, such as pathogen resistance, disease control, sterility, fertility or fruit ripening.

Sequence 475 BP; 163 A; 113 C; 67 G; 132 T; 0 other;

Query Match 100.0%; Score 475; DB 21; Length 475;

Best Local Similarity 100.0%; Pred. No. 1.6e-113;

Matches 475; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATCTTATTCGCGACCTGACTCTCTGTGTGCTGCGGAGGTGCTGCCAAATTTCTGT 60  
 DB 1 GAATCTTATTCGCGACCTGACTCTCTGTGTGCTGCGGAGGTGCTGCCAAATTTCTGT 60  
 QY 61 TCGCGACAACATAGTGTCTGCTGCTGATTCGACAGTCCATATATTTCATGTCAT 120  
 DB 61 TCGCGACAACATAGTGTCTGCTGCTGATTCGACAGTCCATATATTTCATGTCAT 120  
 QY 121 GAGAGACACATGCTAAGTAAATTAATAGCTTAATCCCTTAATACATCAAGAGAT 180  
 DB 121 GAGAGACACATGCTAAGTAAATTAATAGCTTAATCCCTTAATACATCAAGAGAT 180  
 QY 181 GACACATCCAGAGAAAAATCTAATAGTCTTGGGTGGAATGGAACGATAC 240  
 DB 181 GACACATCCAGAGAAAAATCTAATAGTCTTGGGTGGAATGGAACGATAC 240  
 QY 241 CTACATTAATTAACAATTTGCCAAATATAAATAAAGAAAGTTTACATGAGAGTAGT 300  
 DB 241 CTACATTAATTAACAATTTGCCAAATATAAATAAAGAAAGTTTACATGAGAGTAGT 300  
 QY 301 TCTAACATGAGAGACAGTCCAGCAAGCTGACCTATTCGCAAGAGGCTTAGACTTCCA 360  
 DB 301 TCTAACATGAGAGACAGTCCAGCAAGCTGACCTATTCGCAAGAGGCTTAGACTTCCA 360

QY 361 CAATGAGATTAATCCATGAGCTGATGGACACACATCCAAATTAATCCATAAATACCTG 420  
 DB 361 CAATGAGATTAATCCATGAGCTGATGGACACACATCCAAATTAATCCATAAATACCTG 420  
 QY 421 CCCATCCCTCTCTCCAGACTCATCTAATCAAAAAACACACAAACCAATCATG 475  
 DB 421 CCCATCCCTCTCTCCAGACTCATCTAATCAAAAAACACACAAACCAATCATG 475

RESULT 2

ABL70459/C

ID ABL70459 standard; DNA: 16228 BP.

ABL70459;

01-JUL-2002 (first entry)

Chemically treated cell signalling DNA sequence#175.

Cell signalling; cytosine methylation; cell signalling disease;

cancer; tumour; cytostatic; ds.

Unidentified.

W0200202807-A2.

10-JAN-2002.

29-JUN-2001; 2001WO-EP07471.

30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

(EPIC-) EPIGENOMICS AG.

Olex A, Piepenbrock C, Berlin K;

WPI; 2002-154758/20.

Nucleic acid, useful for diagnosis and therapy of diseases associated

PT with cell signalling e.g. cancer, comprises chemically modified genomic

PT sequences of genes associated with cell signalling

PS Claim 1; SEQ ID NO 349; 24pp+sequence listing; English.

The invention relates to a nucleic acid comprising a sequence of at least 18 bases of a segment of chemically pretreated DNA of genes associated with cell signalling. The activity of the modified sequences of the invention may be described as cytostatic. The object of the invention is to provide the chemically modified DNA of genes associated with cell signalling, as well as oligonucleotides and/or PNA-oligomers for detecting cytosine methylations, as well as a method which is particularly suitable for the diagnosis and/or therapy of genetic and chemogenic parameters of genes associated with cell signalling. The chemically modified DNA provided by the invention is useful for diagnosis and therapy of diseases such as solid tumours and cancer. The sequences given in records ABL70111-ABL70626 represent chemically pre-treated genomic DNA's of genes associated with cell signalling. The Note: The sequence data for this patent is not represented in the printed specification, but is based on sequence information supplied by the European Patent Office.

Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;

Query Match 10.4%; Score 49.6; DB 24; Length 16228;

Best Local Similarity 48.3%; Pred. No. 0.0064;

Matches 139; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 131 CATGATTAAGTAAATAGCTTAATCCCTTAATCAATCAACAGATGACATATCA 190  
 DB 9241 CTTTACTATTATTAACCAACTTAATTTTAAAAAATAAATAATATATATA 9182  
 QY 191 CAGAAAAAATCTAATTAATAGCTTTGGGTGATAGAAATGGAACGATACATTAAT 250

```

Db      9181 TAAAAAAATATACATCATTTCTCCACTAAAACTCAAAATCTACATTTATTTAAT 9122
Oy      251 TAAACTTTTGCATATATATATAGAAAGTCTTACCTGAGAGCTAGCTTTAAACATGA 310
Db      9121 ATACCTTTACACACATATATATTTATATAACTCAAAATATATTTTAAACTTTATTTA 9062
Oy      311 AGACTAGTCCACGACGACTGCTTCTATTCACAAAAGGCTTAGACTTTCCACAAATCGAGA 370
Db      9061 AACCAACGCGATATACATCATCTACCTACATCCACAGCTTTAAAAAACCAAAATTAACAAA 9002
Oy      371 TTATCCCATGAGCTGATGACACCATCCAAATTAATCCCTTAATATACC 418
Db      9001 TCACGAATTCATCAAAAAATATAAACCATCTTACCAACATATATAAACCC 8954

RESULT 3
AAS61424/C
ID      AAS61424 standard; DNA; 16228 BP.
AC      AAS61424;
DT      29-JAN-2002 (first entry)
XX      Human gene regulation-associated gene oligonucleotide #379.
DE      Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW      cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW      asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW      renal disease; Preeclampsia; cardiac allograft vascular disease;
KW      colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW      immunostimulant; cardiac; anti-inflammatory; coagulant; antistimetic;
KW      nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
OS      Homo sapiens.
PN      WO200177375-A2.
XX      18-OCT-2001.
PD      06-APR-2001; 2001WO-EP03968.
PF      06-APR-2001; 2000DE-1019058.
PR      07-APR-2001; 2000DE-1019173.
PR      30-JUN-2001; 2000DE-1032529.
PR      01-SEP-2001; 2000DE-1043826.
XX      (EPiG-) EPIGENOMICS AG.
XX      Olek A, Piepenbrock C, Berlin K;
DR      WPI; 2002-017470/02.
XX      New nucleic acid sequences from chemically modified genes associated
PT      with gene regulation, useful for analysing cytosine methylations for
PT      diagnosis and therapy of diseases e.g. severe combined immunodeficiency
XX      disease
XX      Disclosure; SEQ ID NO 387; 26pp; English.
PS      The invention relates to 224 nucleic acid sequences comprising at least
CC      18 bases of a chemically pretreated gene associated with gene regulation
CC      selected from 43 known genes (or complementary sequences). The
CC      chemical pretreatment converts cytosine bases unmethylated at the
CC      5-position to uracil or another base with hybridisation behaviour
CC      dissimilar to cytosine, to enable analysis of cytosine methylations.
CC      The DNA sequences, oligomers (or sets/arrays) and method are
CC      useful in the diagnosis of diseases (or predisposition to diseases)
CC      associated with gene regulation and in therapy of such diseases, by
CC      enabling analysis of the cytosine methylation patterns of such genes,
CC      kits are provided. They are especially useful in diagnosis
CC      and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC      disorders, haemophilia, solid tumours and cancer, Werner syndrome,

```

```

CC      asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC      preclampsia, graft versus-host disease. The present sequence is a
CC      sequence included in the sequence data for this specification and is
CC      associated with the human gene regulation-associated genes.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp://wipo.int/pub/published_pct_sequences
XX      Sequence 16228 BP; 4652 A; 198 C; 3601 G; 7777 T; 0 other;
SQ      Query Match
XX      Best Local Similarity 48.3%; Pred. No. 0.0064;
XX      Matches 139; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
Oy      131 CATGACTTAACTAATTTAGCTTAATCCCTTAACTCAATCAAAAGAGATGACATCA 190
Db      9241 CTTTACTATTTTAAACCAACTTAAATTTTAAAAAATAAACTAAATATATATA 9182
Oy      191 CAGAAAAATTCCTAATAGTCTTGGGTAGAAATGGAAGTGAATGACCTCATTAAT 250
Db      9181 TAAAAAAATATACATCATTTCTCCACTAAAACTCAAAATCTTAAATTTAAT 9122
Oy      251 TACAACTTTGCAATATATATATAGAAAGTCTTAAACATGAGACTAGTCTTAAACATGA 310
Db      9121 ATACTATACACATTAATATTTTATTAATCTTCAAAATTAATTTTAAACCTATTTTA 9062
Oy      311 AGACTAGTCCACGACGACTGCTTCTATTCACAAAAGGCTTAGACTTTCCACAAATCGAGA 370
Db      9061 AACCAACGCGATATACATCATCTACCTACATCCACAGCTTTAAAAAACCAAAATTAACAAA 9002
Oy      371 TTATCCCATGAGCTGATGACACCATCCAAATTAATCCCTTAATATACC 418
Db      9001 TCACGAATTCATCAAAAAATATAAACCATCTTACCAACATATATAAACCC 8954

RESULT 4
ABL34196/C
ID      ABL34196 standard; DNA; 37973 BP.
XX      ABL34196;
AC      26-MAR-2002 (first entry)
DT      Human immune system associated gene SEQ ID NO: 2169.
DE      Human; Immune system disease; cytosine methylation; antiasthmatic;
KW      antiarteriosclerotic; antianaemic; cytosatic; neutropic;
KW      neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW      antirheumatic; antiarthritic; antidiabetic; antipsoriasis;
KW      antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
KW      acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW      neurofibromatosis; Rheumatoid arthritis; psoriasis; bowel disease;
KW      gene; ds.
OS      Homo sapiens.
PN      WO200200928-A2.
XX      03-JAN-2002.
PD      02-JUL-2001; 2001WO-EP07537.
PF      30-JUN-2001; 2000DE-1032529.
PR      01-SEP-2001; 2000DE-1043826.
XX      (EPiG-) EPIGENOMICS AG.
XX      Olek A, Piepenbrock C, Berlin K;
DR      WPI; 2002-130909/17.
XX      Nucleic acid comprising fragment of chemically modified gene, useful
PT

```



KW PNA: cytosine methylation state; SNP: retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner syndrome; developmental disorder;  
 KW psoriasis; Rieger's syndrome; neurologic disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
 KW polyglutamine disorder; solid tumour.  
 XX  
 OS Unidentified.  
 XX  
 XX WO200192565-A2.  
 XX  
 XX 06-DEC-2001.  
 XX  
 XX 06-APR-2001; 2001WO-EP03973.  
 PF  
 XX 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 DR WPI; 2002-09046/12.  
 XX  
 XX New nucleic acids or oligomers, useful for diagnosing or treating  
 PT diseases associated with DNA transcription, e.g., immunological  
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
 PT tumours or cancer.  
 XX  
 PS Claim 1; SEQ ID No 115; 32pp; English.  
 XX  
 XX The invention relates to a nucleic acid, which comprises a segment of the  
 CC chemically pretreated DNA of genes associated with DNA transcription from  
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide  
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 CC to the chemically pretreated DNA of genes associated with DNA  
 CC transcription. The set of oligomer probes are useful for detecting the  
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 CC diagnosing or treating diseases associated with DNA transcription  
 CC (particularly with the methylation status), e.g., adenosine deaminase  
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurologic disorders, neurodegenerative disorders, Waardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 CC or cancer. Sequences ABK28177-ABK28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 XX  
 SO Sequence 15649 BP; 4082 A; 261 C; 3731 G; 7575 T; 0 other:  
 Query Match 10.2%; Score 48.4; DB 24; Length 15649;  
 Best Local Similarity 46.7%; Pred. No. 0.015;  
 Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
 QY 136 CTAAGTAATGCTTAATCCCTTAACCTCAATCAACGAGATGACATCCACAGAA 195  
 DB 11798 CTAACCAACCAACCTACTCTCCCAACGCCCAAAATCAATTTTATATAAAAAACA 11739  
 QY 196 AAAATCTTAATAGCTCTTGCTGTGAAGAAATGGAAGTGAATACCTACATTATACAA 255  
 DB 11738 ACAAAACCTTAACCAACCTAACTATCTTCTTAATAATTAATAAACTAAAAATA 11679

QY 256 CTTTGCNAATTAATTAAGAAAGTCTTACATGAGACTACTTAAATGAGACT 315  
 DB 11678 CATTACATCTTCTTAATTAATTAACCAATTTAAACCATCTCCACCTCAATACCAAAAT 11619  
 QY 316 AGTCACGAACTCGTACCTTATTCACAAAGGCTTAGACTTCCACAAATCGAGATATC 375  
 DB 11618 TTTTCTTCCATTTTTTTTTTTTAAACAAATTTGGCTGATTTCCCAACTAAATACA 11559  
 QY 376 CCATGAGACTGATGACACCATCAATTAATTCCTTAATTAATCTGCCATTTCCCTCTC 435  
 DB 11558 ATAAAGATATCTCACTCACTACCAACCTCCGCTCTAAATTCAAACAATTCCTACTT 11499  
 QY 436 CAGACTCATCTACTCAACAAACACACACA 465  
 DB 11498 CAACCTCTTAATTAATCAATCAATA 11469  
 RESULT 7  
 AAS46626/C  
 ID AAS46626 standard; DNA; 8044 BP.  
 XX  
 XX AAS46626;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Tumour suppressor gene derived chemically modified sequence #348.  
 DE  
 XX Human: tumour suppressor gene; oncogene; antitumour; cytosatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200168912-A2.  
 XX  
 XX 20-SEP-2001.  
 PD  
 XX 15-MAR-2001; 2001WO-EP02955.  
 PF  
 XX 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 DR WPI; 2001-602752/68.  
 XX  
 XX Fragments of chemically modified genes associated with tumour suppressor  
 PT genes and oncogenes, useful in designing primers and probes for  
 PT analysing diseases associated with cytosine methylation state e.g.  
 PT cancer.  
 XX  
 PS Claim 1; SEQ ID No 348; 27pp; English.  
 XX  
 XX The invention relates to a nucleic acid comprising a sequence of 18  
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
 CC bisulphite, of genes associated with tumour suppression and  
 CC oncogenes having a sequence taken from 536 (actually 533 since  
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a  
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
 CC form part of a set of probes for detecting the cytosine methylation state  
 CC and/or single nucleotide polymorphisms and also to be used in an  
 CC array for analysing diseases associated with CpG dinucleotides e.g.  
 CC cancers and tumours. The probes can also be used in a method for  
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
 CC and/or therapy of existing diseases or the predisposition to specific  
 CC diseases, by analysing cytosine methylations. The parameters may be  
 CC compared to another set of genetic and/or epigenetic parameters, the





XX (EPIC-) EPIGENOMICS AG.  
 PA Olek A, Piepenbrock C, Berlin K;  
 PI WPI; 2002-130909/17.  
 DR Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 XX  
 PS Claim 1; SEQ ID NO 542; 32pp + Sequence Listing; German.  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/allergic bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 5856 BP; 1482 A; 177 C; 1495 G; 2702 T; 0 other;

Query Match 9.7%; Score 46; DB 24; Length 5856;  
 Best Local Similarity 46.8%; Pred. No. 0.042;  
 Matches 145; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 160 AAACTCAATCAAAAGAGATGACACATCCACAGAAAAATTCTAATTAGTCTTGGGTG 219  
 DB 5543 AAAATAAACTAAAAAAACCAACCTCAATCTAAATCTATACCCCTTCAATCAA 5484  
 QY 220 TAGAATTTGGAAGTACCTACCTAATTAATTAACCTTTGCAATTAATAAAGAA 279  
 DB 5483 TAAATATTCATTTCCGAAACACACAAACCCACTTCTTCAAAAATATTAATACAA 5424  
 QY 280 AGTTCTACATGAGACACTAGTCTTACATGAAGATGATGACGACACTCGTACCTTATTC 339  
 DB 5423 AAAAAAATAAATTAACATCCACCCACCATTAATCCCACTCTTTTAAAA 5364  
 QY 340 CACAAAGGCTTAGCTTCCAAATGAGATTTCCCATGAGATGATGACACATCCA 399  
 DB 5363 AACCAATCTACTCTTATTAACCAAACTAAATTAATTAACGACATCTGCACTACACAA 5304  
 QY 400 AATTATCCCTAATAATACCTGCGCATTCCTCTCCAGACTCATCTAACCAAAACAA 459  
 DB 5303 AACTTACCTCCCGAATTCAGGCAATCTCTTACTCTAATTAATTAATTAATTA 5244  
 QY 460 CACACACCA 469  
 DB 5243 CAAATACCCA 5234

RESULT 10  
 AAS45299/C  
 ID AAS45299 standard; DNA; 6665 BP.

XX AAS45299;  
 XX 18-DEC-2001 (first entry)

DE Chemically pretreated complementary DNA associated with cell cycle #2.  
 XX  
 KM Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;  
 KM human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 KM graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 KM arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;  
 KM immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;  
 KM PCR primer.

XX Homo sapiens.  
 OS  
 XX WO200168911-A2.

XX 20-SEP-2001.  
 PD  
 XX  
 PF 15-MAR-2001; 2001WO-EP02945.  
 XX  
 XX 15-MAR-2000; 2000DE-1013847.  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2001-602751/68.  
 DR Designing primers and probes for analysing diseases associated with  
 PT cytosine methylation state e.g. arthritis, cancer, aging,  
 PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle.  
 XX  
 PS Claim 1; SEQ ID NO 4; 28pp; English.

XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 CC molecules associated with the cell cycle and specific PCR primers of the  
 CC invention. The sequences are useful for detecting the methylation state  
 CC of all Cpg dinucleotides in a sequence and therefore for analysing  
 CC associated diseases. By analysing cytosine methylations in the pretreated  
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 CC of existing diseases or the predisposition to specific diseases can be  
 CC ascertained. The parameters may be compared to another set of genetic  
 CC and/or epigenetic parameters, the differences serving as basis for  
 CC diagnosis and/or prognosis events which are disadvantageous to patients.  
 CC The sequences of the invention are useful for the diagnosis and therapy  
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 CC aging, glomerular disease, Lewy body disease, arthritis,  
 CC arteriosclerosis, solid tumours and cancers.

XX Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;

Query Match 9.7%; Score 46; DB 22; Length 6665;  
 Best Local Similarity 51.5%; Pred. No. 0.043;  
 Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 88 ATTGACAGTTCATTAATTTATTTCCATGCTGATGAGAGACACATGACTTAAGTATTA 147  
 DB 3168 ATATTAATTACTACACCTTAATTCACACCTTAATAAAACCAAAATTAATTAATTA 3109  
 QY 148 GTTATATCCCTTAATCAATGACAGACATGACACATCCAGAAAAATTCTAATT 207  
 DB 3108 AACCAAAATTCATAAACCAACTAACCAATTAATAAAACCCCACTCTACTATAAA 3049  
 QY 208 AGTCTTGGGTGAGAAATTTGGAAGTGAATACCTTAATTAATTAACACTTTGCAATA 267  
 DB 3048 ATACAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2989  
 QY 268 AATATTAAGAAAGTCTTACATGAA 293  
 DB 2988 AATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2963

RESULT 11  
 ABL32083/C  
 ID ABL32083 standard; DNA; 6665 BP.

XX ABL32083;  
 XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 56.  
 XX  
 KM Human; immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antihaemic; cytostatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiallergic; antidiabetic; antiparasitic;  
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 OS - Homo sapiens.  
 XX WO200200928-A2.  
 PN  
 XX  
 XX  
 XX 03-JAN-2002.  
 PD  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 PF  
 XX  
 XX 30-JUN-2000; 2000DE-1032529.  
 PR  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIDENOMICS AG.  
 PA  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX  
 XX WPI; 2002-130909/17.  
 DR  
 XX  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation  
 PS  
 PS Claim 1; SEQ ID NO 56; 32pp + Sequence Listing; German.  
 XX  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 XX Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;  
 SQ  
 Query Match 9.7%; Score 46; DB 24; Length 6665;  
 Best Local Similarity 51.5%; Pred. No. 0.043;  
 Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
 QY 88 ATTGACAGTTCATTAATTTCCATGTCATGAGAGAGACATGACTAAAGTAATTA 147  
 DB 3168 ATATATACTACACACCTATATTCACACACTTAATAAAACCAAAATTAATTAATCAATTA 3109  
 QY 148 GCTTAATCCCTAAACTCAATACAAAGAGATCAGATCCACAGAAAAATTTCTAATT 207  
 DB 3108 AAACCAAAATTTCAAAACCAACTAACCAACATTAATAAACCCTCTCTACTACTAAA 3049  
 QY 208 AGCTTTGCGTGTGAATGGAATGGAATACCTACATTAATTAACAACCTTTGCAATA 267  
 DB 3048 ATACAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2989  
 QY 268 AAATATAAGAAAGTTCTACATGAA 293  
 DB 2988 AAATATAAGAAAGTTCTACATGAA 2963  
 RESULT 12  
 ABR28130/C  
 ID ABR28130 standard; DNA; 6665 BP.  
 XX  
 XX ABR28130;  
 AC  
 XX  
 XX 23-APR-2002 (first entry)  
 DT  
 XX  
 XX DNA transcription associated complementary genomic DNA #2.  
 DE  
 XX

KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner syndrome; developmental disorder;  
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;  
 KW polyglutamine disorder; solid tumour.  
 XX  
 XX  
 OS unidentified.  
 XX  
 XX  
 XX WO200192565-A2.  
 PN  
 XX  
 XX 06-DEC-2001.  
 PD  
 XX  
 XX 06-APR-2001; 2001WO-EP03973.  
 PF  
 XX  
 XX 06-APR-2000; 2000DE-1019058.  
 PR  
 XX 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIC-) EPIDENOMICS AG.  
 PA  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 PI  
 XX  
 XX WPI; 2002-09046/12.  
 DR  
 XX  
 XX  
 PT New nucleic acids or oligomers, useful for diagnosing or treating  
 PT diseases associated with DNA transcription, e.g. immunological  
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
 PT tumours or cancer  
 PS  
 PS Claim 1; SEQ ID NO 4; 32pp; English.  
 XX  
 XX The invention relates to a nucleic acid, which comprises a segment of the  
 CC chemically pretreated DNA of genes associated with DNA transcription from  
 CC one of 346 sequences, and an oligomer. In particular an oligonucleotide  
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 CC to the chemically pretreated DNA of genes associated with DNA  
 CC transcription. The set of oligomer probes are useful for detecting the  
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 CC diagnosing or treating diseases associated with DNA transcription  
 CC (particularly with the methylation status), e.g. adenosine deaminase  
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurological disorders, neurodegenerative disorders, Waardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart  
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 CC or cancer. Sequences ABR28127-ABR28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 XX  
 SQ Sequence 6665 BP; 1830 A; 86 C; 1243 G; 3506 T; 0 other;  
 Query Match 9.7%; Score 46; DB 24; Length 6665;  
 Best Local Similarity 51.5%; Pred. No. 0.043;  
 Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
 QY 88 ATTGACAGTTCATTAATTTCCATGTCATGAGAGAGACATGACTAAAGTAATTA 147  
 DB 3168 ATATATACTACACACCTATATTCACACACTTAATAAAACCAAAATTAATTAATCAATTA 3109  
 QY 148 GCTTAATCCCTAAACTCAATACAAAGAGATGAGATCCACAGAAAAATTTCTAATT 207  
 DB 3108 AAACCAAAATTTCAAAACCAACTAACCAACATTAATAAACCCTCTCTACTACTAAA 3049

OY		208	AGCCTTGGGTGAGAAATGGAACCATGTAATCCACTTTAATTACAACCTTTGCCAATA	267
Dd		3048	ATTCAAAAATTAATTAATCACTTTAATAATACCCTATTAATCTCAACTTAAAACCTA	298
OY		268	AAATATPAAGANAAGTTCTACATGAA	293
Dd		2988	AAATAAAAAACCTCACTTAACCTAAA	2963
 RESULT 13 ID AAS61101/c DB AAS61101 standard; DNA; 12592 BP.				
XX	AC	AAS61101;		
XX	DT	29-JAN-2002 (first entry)		
XX	DE	Human gene regulation-associated gene oligonucleotide #56.		
XX	KW	Human; Gene regulation-associated gene; severe combined immunodeficiency; cardiac damage; inflammatory response; Haemophilia; Werner syndrome; asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome; renal disease; Preecclampsia; cardiac allograft vascular disease; colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour; immunostimulant; cardiant; antiinflammatory; coagulant; antisthmatic; nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic. OS Homo sapiens.		
XX	PN	WO200177375-A2.		
XX	PD	18-OCT-2001.		
XX	PF	06-APR-2001; 2001WC-EP03968.		
XX	PR	06-APR-2000; 2000DE-1019058.		
XX	PR	07-APR-2000; 2000DE-1019173.		
XX	PR	30-JUN-2000; 2000DE-1032529.		
XX	PR	01-SEP-2000; 2000DE-1043826.		
XX	PT	(EPIg-) EPIGENOMICS AG.		
XX	PI	Olek A, Piepenbrock C, Berlin K; MPI; 2002-017470/02.		
XX	PT	New nucleic acid sequences from chemically modified genes associated with gene regulation, useful for analysing cytosine methylations for diagnosis and therapy of diseases e.g. severe combined immunodeficiency disease -		
XX	PS	Claim 1: SEQ ID No 57; 26pp; English.		
XX	CC	The invention relates to 224 nucleic acid sequences comprising at least 18 bases of a chemically pretreated gene associated with gene regulation selected from 43 known genes (or complementary sequences). The chemical pretreatment converts cytosine bases unmethylated at the 5-position to uracil or another base with hybridisation behaviour dissimilar to cytosine, to enable analysis of cytosine methylations. The DNA sequences, oligomers (or sets/arrays) and method are useful in the diagnosis of diseases (or predisposition to diseases) associated with gene regulation and in therapy of such diseases, by enabling analysis of the cytosine methylation patterns of such genes, kits are provided. They are especially useful in diagnosis and therapy of e.g. severe combined immunodeficiency disease, cardiac disorders, haemophilia, solid tumours and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, pre-eclampsia, graft versus-host disease. The present sequence is a sequence included in the sequence data for this specification and is associated with the human gene regulation-associated genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic		

CC	format directly from WIFO at
CC	ftp.wifo.int/pub/published_pcf_sequences
xx	
50	Sequence 12592 BP; 3783 A; 195 C; 2736 G; 5878 T; 0 other;
	Query Match 9.6%; Score 45.6; DB 24; Length 12592;
	Best Local Similarity 49.7%; Pred. No. 0.065;
	Matches 144; Conservative 0; Mismatches 144; Indels 2; Gaps 1;
Qy	182 ACAACATCCACGAGAAAAATTCTAATTACTCTTTGCGGTGAGAAATGGAAACTGAATACC 241
Db	9190 ACTAATTTTAAAAATAATTAACCCACACACTTAAATATACAAATTCCTCCACACTAAC 9131
Qy	242 TACATTAATTTACACTTTTGGCAATTAATTAAGAGAGTTCTACACTGAAGACT--AG 299
Db	9130 TATCTTAAATATATAATATATATTTAAATATTAACAAACCTCTATTAATTTACTTAA 9071
Qy	300 TTCTAACATGGAAGCTAGTGTGACGACACTCGTACCTTAATTCACAAAGGCTTAGACTTCC 359
Db	9070 CTTAATAATATACTACTTAATTTTCTCTTTTAAAAACAATAATCTCAGCTCTATCA 9011
Qy	360 ACAATGAGATTATCCCATGCACTGATGGACACATCCAAATTTATCCCTTAATAACT 419
Db	9010 CCAACACTAAATTAACAATTAACGCGATCTCGACTACTACAAACCTCCACCTCTAAATCA 8951
Qy	420 GCCCATTCCTCCCTCCGAGACTCATCTTACGCAAAAAACACACAAACCA 469
Db	8950 CGGCATTTCTCTACTCACTCACTCTCTTAATTAACATACTTAACATACTTACCTA 8901
RESULT 14	
ABK33949/C	
ID	ABK33949 standard; DNA; 18997 BP.
xx	
AC	ABK33949;
xx	
DT	18-JUN-2002 (first entry)
xx	
DE	Human DNA for staging of Astrocytomas, complement, #16.
xx	
KW	Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CPG;
KW	disulphide; brain tissue; MALDI; ESI; electron spray mass spectrometry;
KW	matrix assisted laser desorption/ionization mass spectrometry.
xx	
OS	Homo sapiens.
xx	
PN	MO200202808-A2.
xx	
PD	10-JAN-2002.
xx	
PF	02-JUL-2001; 2001WO-EP07538.
xx	
PR	30-JUN-2000; 2000DE-1032529.
xx	
PR	01-SEP-2000; 2000DE-1043826.
xx	
PA	(EPIC-) EPIGENOMICS AG.
xx	
PI	Olek A, Piepenbrock C, Berlin K;
xx	
DR	WPI: 2002-171649/22.
xx	
PT	Novel chemically modified genomic DNA sequences, useful in the
PT	characterisation, classification, differentiation, grading, staging,
PT	treatment and/or diagnosis of astrocytomas or predisposition to
PT	astrocytomas
xx	
PS	Claim 1; SEQ ID No 32; 37bp; English.
xx	
CC	The invention relates to a nucleic acid comprising a sequence (I) of at
CC	least 18 bases in length of a segment of chemically pretreated genomic
CC	DNA which has any one of the sequences of (ABK33919-ABK34032) or its
CC	complement. Also included are an oligonucleotide or peptide nucleic
CC	acid (or set thereof) of at least 9 nucleotides which hybridises to (I),

CC primerisor (1), probes for detecting cytosine methylation or single-  
CC nucleotide polymorphisms (SNP) in (1), an array of oligomers  
CC or peptide nucleic acids for analysing diseases associated with the  
CC methylation states of the CpG dinucleotides of (1). The array is useful  
CC for determining genetic and/or epigenetic parameters, classification,  
CC differentiation, grading, staging, treatment and/or diagnosis of  
CC astrocytomas, or the predisposition to astrocytomas by analysing cytosine  
CC methylations, involves obtaining a biological sample containing genomic  
CC DNA, extracting the genomic DNA, converting cytosine bases which are  
CC unmethylated at the 5-position, in the genomic DNA sample, to uracil or  
CC another base which is dissimilar to cytosine in terms of hybridisation  
CC behaviour, by chemical treatment and amplifying chemically pre-treated  
CC genomic DNA fragments using the array and a polymerase, where the  
CC amplicates carry a detectable label. The method further involves  
CC identifying methylation status of one or more cytosine positions, and  
CC analysing methylation status of the cytosine positions by reference to  
CC one or more data sets. The genomic DNA is chemically treated by using a  
CC bisulphite, hydrogen sulphide or disulphite. The amplification  
CC step amplifies DNA which is of particular interest in astrocytoma or  
CC brain tissue, based on the specific genomic methylation status of brain  
CC tissues, as opposed to background DNA. The amplicates carry a  
CC fluorescent label or radionuclide. Optionally, the labels of the  
CC amplicates are detachable molecule fragments having a typical mass  
CC which are detected in a mass spectrometer. The fragments of chemically  
CC pre-treated genomic DNA to be amplified, have a single positive or  
CC negative charge for a better detectability in the mass spectrometer.  
CC Preferably, the amplicates or fragments of the amplicates are  
CC detected by matrix assisted laser desorption/ionization mass spectrometry  
CC (MALDI) or using electron spray mass spectrometry (ESI). The  
CC present sequence is one of the chemically pre-treated reference DNA  
CC samples of the invention.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 18997 BP; 6146 A; 299 C; 4160 G; 8392 T; 0 other;

Query Match	9.68;	Score 45.4;	DB 24;	Length 18997
Best Local Similarity	40.08;			
Best Global Similarity	40.08;			

Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY	99	CCAAATATATTTTCCATGTGATGAGAGAACACATGACCTAAAGTAATTTAGCTTAATCCCC	158
Db	11885	CAAAACCTATTTTAAACCCTGTATACCTAAATTTAAATTAACAACTAAACCTATTC	11826
QY	159	TAAACGCTATACCAAGAGATGACACACGACGAAAATTTCTAATAGCTTTGGCT	218
Db	11825	TCCTCCCAATTAACAACTAATTAACCTATCATTTAAAAAAAACCTAATTAACAACTATTT	11766
QY	219	GTCGAATTTGGAACGGAATCTACTACATTTAATTACCACTTTTGCAATTAATATTAAGA	278
Db	11765	TAAATATTAACCAATTAATTAATTAATTTTCCAAATCTATTAATTAACCAATTAATA	11706
QY	279	AACTTCTACATGACGACTAGTTCTTACATGAAAGACTAGTCCACGAACTGTAACCTATT	338
Db	11705	AACTATTAATAATTAACCACTTCCAAAAAATAATTCATATTAATTTCTTAATTAC	11648
QY	339	CCACAAA 345	
Db	11645	ACAAAA 11639	

```

RESULT 15
ABL32571/c
ID ABL32571 standard; DNA; 18997 BP.
vv

```

AC ABL32571;

DT 26-MAR-2002 (first entry)  
 YY

DE	Human immune system associated gene	SEQ ID NO: 544
XY		

KM Human immune system diseases, cytosine methylation, antistomatocytic  
KM antileptocystic, antianemic, cytosolic, neurologic  
KM neuroprotective, anti-HIV, anticonvulsant, ophthalmologic,  
KM antihemorrhagic, antithrombotic, antiparasitic, antiparasitic,  
KM antineoplastic, cancer, eye diseases, arteriosclerosis, anaemia,  
KM acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy,  
KM neurofibromatosis, rheumatoid arthritis, psoriasis, bowel disease,  
KM gene, ds.

OS Homo sapiens.  
YY

PN WO200200928-A2  
XX

PD 03-JAN-2002.  
XX

02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.

XX  
PA (EPIC-) EPIGENOMICS AG

PI Olek A, Piepenbrock C, Berlin K

AA  
DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful

cytosine methylation

PS Claim 1; SEQ ID NO 544; 32pp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/allergic bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 1697 BP; 6146 A; 299 C; 4160 G; 8392 T; 0 other;

Query Match	9.68;	Score 45.4;	DB 24;	Length 18997;
Post Local Similarity	40.08;	Score 40.08;	DB 24;	Length 18997;

```
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
```

QY	99	CCAAATATTATTTTCCCATGTCATGAGGAAGACACATATGCTAATAACTATTATAGCTTAATCCCC	158
Dd	11885	CAAAACCTATTTTTAAAAACCCTGTATTAACCTAAATTTAAATAAACAAACTAAACCTACTTTC	11826
QY	159	TAAACCTCAATACAAAAGSAGATGACAATCCACAGAAAATAATCTATTAGCTTTGGCT	218
Dd	11825	TCCCTGCATATAAAAACTAAATTAAGTTATTCATTATTAATAAAAAAACCTAATATAAAAACAATAT	11766
QY	219	GTAGAAATTTGGAACACGAACTACTACATTATATTCACACTTTGCCAATAATAATATAAA	278
Dd	11765	TAAATTAATAAAAACCAATRAAAAAAAAAATTAAATTTCCAAATCTATTAATAATACAAATATATA	11706
QY	279	AAGTTTAACTAGTAAGAAGTACTAGTCTAACATGAAGAGTACTAGCCAGAACGTAACCTATT	338
Dd	11705	AACCTATAAAAATTAATAAACCACTCCAAAAAATAAAAAAATTCATTAATATCTTAATATACC	11646
QY	339	CCACAAA 345	
Dd	11645	ACAAAAA 11639	

Search completed: July 20, 2003, 04:53:23  
Job time : 193 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 03:30:14 ; Search time 44 Seconds  
(without alignments)  
3310.715 Million cell updates/sec

Title: US-09-719-002-1

Sequence: 1 gaattcttatctgcagcctga.....acacacacacacacatcatg 475

Scoring table: IDENTITY\_NDC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued\_Patents\_NA: \*  
1: /cgn2\_6/prodata/1/lna/5A.COMB.seq: \*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq: \*  
3: /cgn2\_6/prodata/1/lna/6A.COMB.seq: \*  
4: /cgn2\_6/prodata/1/lna/6B.COMB.seq: \*  
5: /cgn2\_6/prodata/1/lna/PTOS.COMB.seq: \*  
6: /cgn2\_6/prodata/1/lna/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	8.0	1263	US-09-302-681-4	Sequence 4, Appli
2	36.8	7.7	6243	US-09-056-075-1	Sequence 1, Appli
3	36.4	7.7	4093	US-07-861-458C-5	Sequence 1, Appli
4	35	7.4	3815	US-06-936-165A-196	Sequence 196, App
5	34.4	7.2	457	US-09-643-597-306	Sequence 306, App
6	33.8	7.1	24595	5428147-1	Patent No. 5428147
7	33.6	7.1	4673	US-07-638-431-1	Sequence 1, Appli
8	33.6	7.1	4673	PCT-US92-00018-1	Sequence 1, Appli
9	33.6	7.1	9412	US-08-955-138-1	Sequence 1, Appli
10	33.6	7.1	168575	US-09-426-290-1	Sequence 1, Appli
11	33.2	7.0	1253	US-08-591-628-1	Sequence 1, Appli
12	33.2	7.0	8700	US-08-392-625-16	Sequence 16, Appli
13	33.2	7.0	8700	US-08-466-961A-16	Sequence 16, Appli
14	33.2	7.0	8700	US-08-645-193B-18	Sequence 18, Appli
15	33	6.9	357	US-08-666-405-12	Sequence 12, Appli
16	33	6.9	575	US-08-569-166-29	Sequence 29, Appli
17	33	6.9	882	US-08-666-405-27	Sequence 27, Appli
18	32.8	6.9	246240	US-08-724-394A-20	Sequence 20, Appli
19	32.8	6.9	246240	US-08-724-394A-21	Sequence 21, Appli
20	32.8	6.9	246240	US-08-724-394A-22	Sequence 22, Appli
21	32.6	6.9	1001	US-09-641-638-558	Sequence 558, App
22	32.6	6.9	1001	US-09-641-638-559	Sequence 559, App
23	32.6	6.9	12793	US-09-004-838-124	Sequence 124, App
24	32.6	6.9	15062	US-09-004-838-189	Sequence 189, Appli
25	32.6	6.9	19124	US-08-487-828B-13	Sequence 13, Appli
26	32.4	6.8	359	US-08-687-080-65	Sequence 65, Appli
27	32.4	6.8	5609	US-09-313-677-14	Sequence 14, Appli

28	32.4	6.8	5609	US-09-313-677-15	Sequence 15, Appli
29	32.4	6.8	7026	US-09-313-677-20	Sequence 20, Appli
30	32.4	6.8	7344	US-09-313-677-16	Sequence 16, Appli
31	32.4	6.8	98844	US-09-791-211-10	Sequence 10, Appli
32	32	6.7	600	US-09-385-982-482	Sequence 482, App
33	32	6.7	733	US-09-385-982-541	Sequence 541, App
34	32	6.7	2639	US-09-374-135-1	Sequence 1, Appli
35	31.8	6.7	246240	US-08-724-394A-20	Sequence 20, Appli
36	31.8	6.7	246240	US-08-724-394A-21	Sequence 21, Appli
37	31.8	6.7	246240	US-08-724-394A-22	Sequence 22, Appli
38	31.6	6.7	658	US-08-996-416-595	Sequence 595, App
39	31.6	6.7	731	US-08-451-405A-2	Sequence 2, Appli
40	31.6	6.7	3558	US-09-134-001C-2439	Sequence 2439, Ap
41	31.6	6.7	13737	US-09-538-414-10	Sequence 10, Appli
42	31.4	6.6	568	US-08-642-214D-23	Sequence 23, Appli
43	31.4	6.6	568	US-08-952-014C-23	Sequence 23, Appli
44	31.4	6.6	2539	US-08-749-522-3	Sequence 3, Appli
45	31.4	6.6	5099	US-09-610-040-5	Sequence 5, Appli

## ALIGNMENTS

```
RESULT 1
US-09-302-681-4
Sequence 4, Application US/09302681
Patent No. 6441149
GENERAL INFORMATION:
APPLICANT: Hermsstadt, Corinna
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Cleverger, William
APPLICANT: Fahy, Eoin F.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON
TITLE OF INVENTION: QUANTIFICATION OF EXTRAMITOCHONDRIAL DNA
FILE REFERENCE: 660088.416C1
CURRENT APPLICATION NUMBER: US/09/302,681
CURRENT FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 108
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 1263
TYPE: DNA
ORGANISM: Homo sapien
US-09-302-681-4

Query Match      8.0%   Score 38;   DB 4;   Length 1263;
Best Local Similarity 57.6%   Pred. No. 0.12;
Matches 68;   Conservative 0;   Mismatches 50;   Indels 0;   Gaps 0;

Cy 102 ATATTATTTCATGTCATGAGAGACACATGCTAAGTAATAGCTTAATCCCTAA 161
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 803 ATGATTGTGACATATATGTCACATGATGCAATGAAATTAATTAAGGTATATCCCAAT 862

Cy 162 AACCATTAACAAGCAATGACATCCACAGAAAAATTCATTAATGCTTTGGGTG 219
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 863 AAATTAACAAGCAATTTTAAAGAGAGAGAGAGAGAGATGGGCTTTGGGTG 920

RESULT 2
US-09-056-075-1
Sequence 1, Application US/09056075
Patent No. 5953568
GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marile
APPLICANT: Rood, Julian
TITLE OF INVENTION: Expression System for Clostridium
TITLE OF INVENTION: Species
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Charles & Brady
STREET: 1 South Pinckney Street
```

CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
CURRENT APPLICATION DATA:  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/09/056,075  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note="RP4 origin of DNA transfer (orit) from  
US-09-056-075-1"

Query Match  
Best Local Similarity 7.7%; Score 36.8; DB 2; Length 6243;  
Matches 95; Conservative 0; Mismatches 97; Indels 0; Gaps 0;  
QY 122 AGAGAGACATGACTTAAGTATAGCTTAATCCCTAAACCTCAATGCAAGAGATG 181  
DB 1213 AAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1272  
QY 182 ACACATCCACAGAAAAATCTAATTAGCTTTGGCTTAAGAAATGGAATGCC 241  
DB 1273 AAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1332  
QY 242 TACATTAACTCAACTTTGCAATTAATTAAGAAAGTCTACATGAGAGCTGTT 301  
DB 1333 TATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1392  
QY 302 CTACATGAGAGA 313  
DB 1393 AAAAAATTAATAATA 1404

RESULT 3  
US-07-861-458C-5/C  
Sequence 5, Application US/07861458C  
Patent No. 6232061  
GENERAL INFORMATION:  
APPLICANT: Marchionni, Mark Andrew  
APPLICANT: Johnson, Carl D.  
TITLE OF INVENTION: HOMOLOGY CLONING  
NUMBER OF SEQUENCES: 142  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/861,458C  
FILING DATE: 04/01/92  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 04585/014001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4093  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-861-458C-5

Query Match  
Best Local Similarity 7.7%; Score 36.4; DB 4; Length 4093;  
Matches 76; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 33 GCTGCCAGGTGCTGTCGAAATTTCTGTTGCGCACAACTACTGTCCTGTTGATTG 92  
DB 1432 GCTTGGCGCTGTTCCTGCGATCTTGAGCGCACAGAAACGAGTTCTTAATAA 1373  
QY 93 ACAGTTCATTAATTAATTCATGTCATGAGAGACACATGCTAAAGTAATTA 152  
DB 1372 ATAGTTAAATAAATGGCTTAATAATAATAGTTGAATTAATGCTTAATAATA 1313  
QY 153 ATCCCTAAACTCAATACAA 174  
DB 1312 ATAAATAAAGATATTAGAA 1291

RESULT 4  
US-08-936-165A-196/C  
Sequence 196, Application US/08936165A  
Patent No. 6348582  
GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lometto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: Pol. 6348582el Prokaryotic Polynucleotides,  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

Query Match Similarity	7.2%	Score 34.4	DB 4	Length 457
Best Local Similarity	54.9%	Pred. No. 0.87		
Matches	89	Conservative	0	Mismatches 71; Indels 2; Gaps 1
QY	138	AAAGTATATAGCTTATCCCTAAACTCAATCAACAAGATGACACATCCA--CAGAA	195	
Db	9	AAGGTATATACCTTATATATATTTAAAGACACATGCAACGAAAAACGAATGTGACAGTGC	68	

Query Match	7.13;	Score 33.8;	DB 6;	Length 24595;
Best Local Similarity	51.08;	Pred. No. 4.9;		
Matches 80;	Conservative	0;	Mismatches	77; Indels 0; Gaps 0;
Db	200	TTCTAATTACTCTTTGGCGTGTAGAAATTTGGAACCTGAATCTACATTAACTTACACTTT	259	
Db	23880	TTCCCTTCAAGCTGGCGATATATGAACTGATAGGCGCTCCAAAAAATTTCACCCG	23822	
QY	260	TGCAATTAATTAATTAAGAGTTCTACATGAGACTACTTCTAACATGAGACCTAGTC	319	
Db	23820	ACAAACAACAAATACAGAAATTTAGACATTAATAATGAATTCACACCGCATATTCG	23761	
QY	320	CACGAAGCTGTACTTATTCCACAAAGGCTTAACCT	356	
Db	23760	CCATCAGTCATTCACAAAATCACCAATTAATGAGCTTT	23724	

RESULT 7  
US-07-638-431-1/c  
Sequence 1, Application US/07638431  
Patent No. 5198535  
GENERAL INFORMATION:  
APPLICANT: Hoffman, Stephen L.  
APPLICANT: Charoenvit, Yupin  
APPLICANT: Hedstrom, Richard  
APPLICANT: Khushf, Srisin  
APPLICANT: Rogers IV, William O.

```

      TITLE OF INVENTION: Protective malaria sporozoite surface protein
      TITLE OF INVENTION: Immunogen and gene
      NUMBER OF SEQUENCES: 2
      CORRESPONDENCE ADDRESSES:
      ADDRESSEE: A. David Spevack
      STREET: NMDC Building 1 T-12 National Naval
      CITY: Bethesda
      STATE: MD
      COUNTRY: USA
      ZIP: 20814-5044
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: IBM PC compatible
      SOFTWARE: Patentin Release #1.24
      CURRENT APPLICATION NUMBER: US/07/638,431
      FILING DATE: 19910110
      CLASSIFICATION: 424
      ATTORNEY/AGENT INFORMATION:
      NAME: Spevack, Avram D.
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (301) 295-6759
      TELEFAX: (301) 295-4033
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 4673 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      HYPOTHEetical: N
      ANTI-SENSE: N
      ORIGINAL SOURCE:
      ORGANISM: Plasmodium yoelii
      STRAIN: 17X(NL)
      DEVELOPMENTAL STAGE: erythrocytic stage
      TISSUE TYPE: Blood
      CELL TYPE: erythrocytic stage
      IMMEDIATE SOURCE:
      LIBRARY: Py-lambda gtl1-2-7 kb genomic expression
      CLONE: Py10.1111
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 718..3195
      OTHER INFORMATION:
      US-07-638-431-1

Query Match
Best Local Similarity 48.0%; Pred No. 3.2; Length 4673;
Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 93 ACAGTCCAAATTAATTTCCATGTCATGAGAGAGACATGACTAAATTAAGCTTA 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4427 ACAATATAGAAATATTTCTTCCATATATATGACACAGACAAATATGATGAATA 4368
QY 153 ATCCCTAAACTCATCAACGAGATGACACATCCACAAAAATTCATTAAGCT 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4367 TATATACATCTACTCAAAATTAATTAATTTCAAAATTAAGAAATTAAGAAATTA 4308
QY 213 TTGCGTGTAGAAATTTGAAATGAAATACCTACATTAATTCACAGTTTGGCAATTAATA 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4307 ATAAATTTATATGCTCCAAACAAAAATGACACATGCACTACAAATTAATGAAGAAAT 4248
QY 273 TAAAGAAAGTCTTAACATGA 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4247 GAAATTAATTTGAAATGA 4228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
PCT-US92-00018-1/c
Sequence 1, Application PC/TUS9200018

      GENERAL INFORMATION:
      APPLICANT: Hoffman, Stephen L.
      APPLICANT: Chareonvity, Yupin
      APPLICANT: Hedstrom, Richard
      APPLICANT: Khushnith, Srisin
      APPLICANT: Rogers IV, William O.
      TITLE OF INVENTION: Protective malaria sporozoite surface protein
      NUMBER OF SEQUENCES: 2
      CORRESPONDENCE ADDRESSES:
      ADDRESSEE: A. David Spevack
      STREET: NMDC Building 1 T-12 National Naval
      CITY: Bethesda
      STATE: MD
      COUNTRY: USA
      ZIP: 20814-5044
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      OPERATING SYSTEM: IBM PC compatible
      SOFTWARE: Patentin Release #1.24
      CURRENT APPLICATION NUMBER: PCT/US92/00018
      FILING DATE: 19920103
      CLASSIFICATION: 424
      ATTORNEY/AGENT INFORMATION:
      NAME: Spevack, Avram D.
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (301) 295-6759
      TELEFAX: (301) 295-4033
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 4673 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      HYPOTHEtical: N
      ANTI-SENSE: N
      ORIGINAL SOURCE:
      ORGANISM: Plasmodium yoelii
      STRAIN: 17X(NL)
      DEVELOPMENTAL STAGE: erythrocytic stage
      TISSUE TYPE: Blood
      CELL TYPE: erythrocytic stage
      IMMEDIATE SOURCE:
      LIBRARY: Py-lambda gtl1-2-7 kb genomic expression
      CLONE: Py10.1111
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 718..3195
      OTHER INFORMATION:
      PCT-US92-00018-1

Query Match
Best Local Similarity 48.0%; Pred No. 3.2; Length 4673;
Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 93 ACAGTCCAAATTAATTTCCATGTCATGAGAGAGACATGACTAAATTAAGCTTA 152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4427 ACAATATAGAAATATTTCTTCCATATATATGACACAGACAAATATGATGAATA 4368
QY 153 ATCCCTAAACTCATCAACGAGATGACACATCCACAAAAATTCATTAAGCT 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4367 TATATACATCTACTCAAAATTAATTAATTTCAAAATTAAGAAATTAAGAAATTA 4308
QY 213 TTGCGTGTAGAAATTTGAAATGAAATACCTACATTAATTCACAGTTTGGCAATTAATA 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4307 ATAAATTTATGCTCCAAACAAAAATGACACATGCACTACAAATTAATGAAGAAAT 4248
QY 273 TAAAGAAAGTCTTAACATGA 292
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4247 GAAATTAATTTGAAATGA 4228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```

Db      4247 GAAATAAATGGGAAAATGA 4228

RESULT 9
US-08-955-138-1/c
; Sequence 1, Application US/08955138A
; Patent No. 5977435
; GENERAL INFORMATION:
; APPLICANT: Lefebvre, Daniel D.
; APPLICANT: Gelliaty, Kevin S.
; TITLE OF INVENTION: PLANT PHOSPHATASES
; FILE REFERENCE: PPL97-01
; CURRENT APPLICATION NUMBER: US/08/955,138A
; CURRENT FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 119
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO. 1
; LENGTH: 9412
; TYPE: DNA
; ORGANISM: SOLANUM TUBEROSUM
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(17)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-08-955-138-1

Query Match          7.1%; Score 33.6; DB 2; Length 9412;
Best Local Similarity 51.5%; Pred. NO. 4;
Matches 101; Conservative 0; Mismatches 94; Indels 1; Gaps 1;

QY    142 TAATTAGCTTAATCCCTTAAACTACATAACAACGAGATGACACATCCACACAAAAAATT 201
      ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2535 TAAGCAGCTTAATCTTTTCATTAAGACAATGAAATAATAGACAGTACAGATTGGAT 2476

QY    202 CTAATT-ASTCTTTGCGTGAGAAATTGGAACGTGAATACCCTACATTAATACAACTTT 260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2475 CAATCTCTCAACATATGAGCCAGTGAAGAAGCAATGCACAAAAAATTTACAGCTTT 2416

QY    261 GCATAAATAATATAAGAAAGTTCTAACATGAAGACTAGTCTTAAACATGAAGACTAGTCC 320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2415 GCATATATATAGTAGAAAAATGAATAAGAAAAAATAAAGCTTATCTATCTAGTCC 2356

QY    321 ACGRACCTGTACCTTA 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    2355 ATCACTGCTGCATCA 2340

RESULT 10
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Bergling Ran Olafsdottir
; APPLICANT: Jeffrey Guicher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO. 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)

NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match          7.1%; Score 33.6; DB 4; Length 168575;
Best Local Similarity 49.4%; Pred. NO. 11;
Matches 87; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY    204 AATTAGTCTTTGGTGTTAGAAATGGAAACTCAATACCTACATTAATTACAACTTTTGCA 263
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    53841 AAATAAATTTCTCTTTATTGTTATTAGTGAATGTACAATTTATTGATCAATATAAATGAA 53782

QY    264 AATAAATAATAAGAAAGTTCTTAACATGAAGACTAGTCTTAACATGAAGACTAGTCCAG 323
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    53781 ATAATAATTTAGAAATTTATTCCAACTTTGCTCAATGTCATGCAACGAGACTCCATCTCA 53722

QY    324 AACTCGTACCTTATTCCACAAAAGGTTTAGACITTTCCACAAAATCCGAGATTATCCCAT 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    53721 ATAATCATGAATTTCCAAAGATTTCTTAGATTATTAAGATGCAAAAGATTCTTAGAT 53666

RESULT 11
US-08-591-629-1/c
; Sequence 1, Application US/08591629
; Patent No. 5993808
; GENERAL INFORMATION:
; APPLICANT: MELCHERS, Leo Sjoerd
; APPLICANT: APOTHEKER-DE GROOT, Marion
; APPLICANT: BOL, John Ferdinand
; APPLICANT: CORNELISSEN, Bernardus Johannes Clemens
; APPLICANT: LINHORST, Rubertus Josephus Maria
; APPLICANT: PINSTEIN, Anne Silene
; APPLICANT: SELA-BUURLAGE, Marianne Beatrix
; TITLE OF INVENTION: Plant chitinases, DNA coding therefor and
; METHOD OF INVENTION: plants containing same
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ladass & Parry
; STREET: 26 West 61st Street
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10023-7604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM PC 4.86 SX 50 Mhz
; OPERATING SYSTEM: DOS 6.20
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,629
; FILING DATE: 15-FEB-96
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/02761
; FILING DATE: 17-AUG-94
; PRIOR APPLICATION DATA: EP 93202425.0
; APPLICATION NUMBER: 17-AUG-93
; ATTORNEY/AGENT INFORMATION:
; NAME: MASS, CLIFFORD J.
; REGISTRATION NUMBER: 30,086
; REFERENCE/DOCKET NUMBER: U-010627-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 708-1800
; TELEFAX: (212) 246-8959
; TELEX: 233288
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; - HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Nicotiana tabacum
; STRAIN: Samsun NN
; DEVELOPMENTAL STAGE: TMV-induced
; IMMEDIATE SOURCE:
; CLONE: Cluster-A
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14..1126
; OTHER INFORMATION: /partial
US-08-591-629-1

```

```

Query Match 7.0%; Score 33.2; DB 2; Length 1253;
Best Local Similarity 53.0%; Pred. No. 2.6;
Matches 71; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 54 TTTCTGTTGCCACACACATAGTCTGCTTGGTTGATTGACAGTCCCAATAATTATTTCC 1113
DB 1232 TTTTGGAGACCAAAATTATTGTTTATTTATGTCATGTCACATATCTGTGCC 1173
QY 114 ATGTCATGAGAGACGACATGACTAAAGTAAATAGCTTAATCCCTAAACTCAATACAA 173
DB 1172 TTATTATAAGTAGTATATCTTGACACACACAAATTACGTAATCCCATCTCTCTGAG 1113
QY 174 ACGAGATGACACAT 187
DB 1112 ATGACACTCCCAT 1099

```

```

RESULT 12
US-08-392-625-16/c
; Sequence 16, Application US/08392625
; Patent No. 5837485
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5837485bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Germar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic Process For The Preparation
; OF CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,625
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,791
; FILING DATE: 30-APR-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8700 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-392-625-16
Query Match 7.0%; Score 33.2; DB 2; Length 8700;
Best Local Similarity 47.6%; Pred. No. 5;
Matches 98; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 128 GCACATGACTAAAGTAATTAGCTTAATCCCTPAAACTCAATACAAACGAGATGACACAT 187
DB 8306 GAAGTTCACATAATGAAGAATTAAAGTTATATATAGTGTGAATATATGATATGCAAAACTT 8247
QY 188 CCACAGAAAAAATCTTAATTAGTCTTTGGTGTAGAAATGGAAACTGAATACCTACATT 247
DB 8246 TTAAGATCCCATTAACAAAAAATATCGAGTTAACTTATAAAATTTCTGTGACTACATA 8187
QY 248 AATTACAACATTTGGCAATPAAATATAAAGAAAGTTCTTAACATGAAGACTAGTTCTTAACA 307
DB 8186 CAGCACAAATTAAACTAGCAAAATTAATTTCACTATAAAGAGTATAAAAAACG 8127
QY 308 TGAAGACTAGTCCAGCACTCTGACC 333
DB 8126 TCAAGTTTGTAAATCCTACCTGTTCC 8101

```

```

RESULT 13
US-08-466-961A-16/c
; Sequence 15, Application US/08466961A
; Patent No. 5843709
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5843709bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Germar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic process for the preparation of
; chemical compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,961A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/392,625  
FILING DATE: 22-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/676,791  
FILING DATE: 30-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/784,234  
FILING DATE: 31-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.0980004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 16:  
LENGTH: 8700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-466-961A-16

Query Match 7.08; Score 33.2; DB 2; Length 8700;  
Best Local Similarity 47.6%; Pred. No. 5;  
Matches 98; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 128 GCACATGACGTAAGTAATTAGCTTAATCCCTAAACCTCAATCAAAACGAGATGACACAT 187  
DB 8306 GAAGTTCTACAAATGAAGAATTAAGTTATTATAGTGTGATATGATTAATGCAAAACTT 8247  
QY 188 CCACAGAAAAATCTTAATTAGCTTTGGCGTAGAAATGGAACTGAATGATCTACATT 247  
DB 8246 TTAAGAATCCATTAAACAAAAAATATCGAGTTAACTTATAAAATTCCTGAACTACATA 8187  
QY 248 AATTACAACTTTCCCAATATAAAGAAAGTTCTTAACATGAAGACTAGTTCTAAACA 307  
DB 8186 CAGCACAATTAAGAACTAGCAAAATCAAAATTAATTCATTAAGAGTAATAAAACG 8127  
QY 308 TGAAGACTAGTCCACGACTCGTACC 333  
DB 8126 TCAAGTTGTAATCCCTACCTGTCC 8101

RESULT 14  
US-08-645-193B-18/c  
Sequence 18, Application US/08645193B  
Patent No. 5962253  
GENERAL INFORMATION:  
APPLICANT: Kupke, Thomas  
APPLICANT: Gotz, Friedrich  
APPLICANT: Kempter, Christoph  
APPLICANT: Jung, Gunther  
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides  
TITLE OF INVENTION: Catalyzed by Flavoprotein Epid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/645,193B  
FILING DATE: 13-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.1540000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: cDNA  
US-08-645-193B-18

Query Match 7.0%; Score 33.2; DB 2; Length 8700;  
Best Local Similarity 47.6%; Pred. No. 5;  
Matches 98; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
QY 128 GCACATGACGTAAGTAATTAGCTTAATCCCTAAACCTCAATCAAAACGAGATGACACAT 187  
DB 8306 GAAGTTCTACAAATGAAGAATTAAGTTATTATAGTGTGATATGATTAATGCAAAACTT 8247  
QY 188 CCACAGAAAAATCTTAATTAGCTTTGGCGTAGAAATGGAACTGAATGATCTACATT 247  
DB 8246 TTAAGAATCCATTAAACAAAAAATATCGAGTTAACTTATAAAATTCCTGAACTACATA 8187  
QY 248 AATTACAACTTTCCCAATATAAAGAAAGTTCTTAACATGAAGACTAGTTCTAAACA 307  
DB 8186 CAGCACAATTAAGAACTAGCAAAATCAAAATTAATTCATTAAGAGTAATAAAACG 8127  
QY 308 TGAAGACTAGTCCACGACTCGTACC 333  
DB 8126 TCAAGTTGTAATCCCTACCTGTCC 8101

RESULT 15  
US-08-666-405-12  
Sequence 12, Application US/08666405  
Patent No. 5874220  
GENERAL INFORMATION:  
APPLICANT: FACH, Patrick; GUILLOU,  
APPLICANT: Jean-Pierre; POPOFF, Michel  
TITLE OF INVENTION: PRIMERS FOR THE  
TITLE OF INVENTION: AMPLIFICATION OF GENES CODING FOR THE  
TITLE OF INVENTION: ENTEROTOXIN AND THE LEGITHINASE OF CLOSTRIDIUM  
TITLE OF INVENTION: PERRINGENS AND THEIR APPLICATION TO THE  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIERMAN & MUSERLIAN  
STREET: 600 THIRD AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/666,405  
FILING DATE: 08-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP94/04292  
FILING DATE: 22-DEC-1994  
PRIOR APPLICATION DATA: US 08/172,026  
FILING DATE: 22-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: MUSERLIAN, CHARLES A

Search completed: July 20, 2003, 05:20:04  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 05:19:14 ; Search time 171 Seconds  
(without alignments)  
5730.558 Million cell updates/sec

Title: US-09-719-002-1  
Perfect score: 475  
Sequence: 1 gaattttattgcaccta.....acaacacacacacacatcatg 475

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 1439767 seqs, 1031500376 residues

Total number of hits satisfying chosen parameters: 2879534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*

10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq1:\*

11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq2:\*

12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq3:\*

13: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*

15: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*

17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	48.4	10.2	15649	15	US-10-239-676-103
C 2	46	9.7	6665	15	US-10-239-676-4
C 3	45.4	9.6	18997	15	US-10-172-086-18
C 4	43.2	9.1	11812	15	US-10-239-676-210
C 5	42.8	9.0	513509	12	US-09-754-853A-4
C 6	42.4	8.9	15732	15	US-10-239-676-96
C 7	41.8	8.8	4504	15	US-10-239-676-200
C 8	41	8.6	18997	15	US-10-172-086-17
C 9	40.8	8.6	6158	15	US-10-239-676-24
C 10	40.2	8.5	7195	15	US-10-239-676-30
C 11	39.8	8.4	414	11	US-09-960-352-6528
C 12	39.8	8.4	6050	15	US-10-172-086-50
C 13	39.4	8.3	640681	11	US-09-790-988-1
C 14	39	8.2	6167	15	US-10-239-676-41
C 15	38.8	8.2	864	11	US-10-938-842A-4290
C 16	38.8	8.2	6061	15	US-10-239-676-40

C 17	38.8	8.2	6478	15	US-10-239-676-123
C 18	38.4	8.1	3158	15	US-10-078-929-187
C 19	38.4	8.1	6074	15	US-10-172-086-43
C 20	38.4	8.1	12405	15	US-10-239-676-35
C 21	38.2	8.0	2000	11	US-09-887-576-204
C 22	38	8.0	8842	15	US-10-239-676-72
C 23	37.4	7.9	7432	15	US-10-172-086-3
C 24	37.2	7.8	1347	15	US-10-027-632-212736
C 25	37.2	7.8	7040	15	US-10-172-086-14
C 26	37.2	7.8	15732	15	US-10-239-676-95
C 27	37	7.8	543	15	US-10-027-632-109062
C 28	37	7.8	543	15	US-10-027-632-109063
C 29	37	7.8	1448	8	US-08-781-986A-709
C 30	37	7.8	5834	15	US-10-239-676-135
C 31	37	7.8	6158	15	US-10-239-676-23
C 32	37	7.8	177556	12	US-09-952-213D-6
C 33	36.8	7.7	446	11	US-09-960-352-3400
C 34	36.8	7.7	624	15	US-10-027-632-214763
C 35	36.8	7.7	624	15	US-10-027-632-214764
C 36	36.8	7.7	8996	15	US-10-239-676-211
C 37	36.8	7.7	9539	15	US-10-239-676-52
C 38	36.8	7.7	10286	15	US-10-239-676-13
C 39	36.8	7.7	14849	15	US-10-239-676-121
C 40	36.6	7.7	5690	15	US-10-239-676-74
C 41	36.4	7.7	11036	15	US-10-239-676-118
C 42	36.4	7.7	6621	15	US-10-172-086-9
C 43	36.4	7.7	7089	15	US-10-239-676-68
C 44	36.4	7.7	7588	15	US-10-239-676-191
C 45	36.2	7.6	12465	15	US-10-239-676-31

## ALIGNMENTS

## RESULT 1

US-10-239-676-103/C  
; Sequence 103, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-06  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 103  
; LENGTH: 15649  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-103

Query Match 10.28; Score 48.4; DB 15; Length 15649;

Best Local Similarity 46.7%; Pred. No. 0.047;

Matches 154; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

Oy 136 CTAAGTAATAGCTTAATCCCTAAACATCAATACAAACGAGATGACATCCACAGAA 195

Db 11798 CTAACACCAACCAACTACTCTTCCAACTCCCAAAATCAATTTTTCATAAAAAAACA 11739

QY 196 AARATCTAATTAAGTCTTGGTCTACAGAAATGCAAACTGAATACCTACATTAATTAACAA 255  
Db 11738 ACAAACTTTAAACACATAACTATCTTCTTAAATTAATTAATAAACTAAATAA 11679  
QY 256 CTTTGGCAATATAAAGAAAGTCTTAACATGAAGACTAGTTCTACATGAAGACT 315  
Db 11678 CATTAACCTTTCTAATTTATACACAATTTAAACACATCCCTCAACCTCAAACTCAAAAT 11619  
QY 316 AGTCCAGACTCGTACTTATCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATC 375  
Db 11618 TTTTCTTCCATTTTTTTTTTAAACAAATTTCCGCTCTATTCCGCCAACTAAATACA 11559  
QY 376 CCATGGACTGATGACACCATCAAAATTTATCCCTATATAATACCTTGCCCATCCCTCCCTC 435  
Db 11558 ATAACGATATCTCAACTCACTACACCTCGCCCTCTAAATTTCAACAAATTCCTCTACTT 11499  
QY 436 CAGACTCATCTACTCAAAAACACACACA 465  
Db 11498 CAACCTCTCTAAATAACTTAAACTACAATA 11469

RESULT 2  
US-10-239-676-4/c  
; Sequence 4, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 4  
; LENGTH: 6665  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-4

Query Match 9.7%; Score 46; DB 15; Length 6665;  
Best Local Similarity 51.5%; Pred. No. 0.13;  
Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 88 ATTTGACAGTTCCTCAATTAATTTCCATGTCATGAGAGGACATGACTAAAGTAATTA 147  
Db 3168 ATATAATAACTCACACCTATTAATTTCCACACTCTTAAACCCAAATTAATTAATCAATTA 3109  
QY 148 GCTTAATCCCTTAAATCAATAACAGGAGATGACATCCACAGAAAAATTTCTAAT 207  
Db 3108 AAACCAAAATTTCAAAACCAATTAACCAATTAATAAATTAATTAATTAATTAATTAATTA 3049  
QY 208 AGTCTTTCGCTGTAAGAAATGGAATGAACTGAACTACATTAATTAATTAATTAATTAATTA 267  
Db 3048 ATACAAAAATTAATAACITTAATAATAATTAATTAATTAATTAATTAATTAATTAATTA 2989  
QY 268 AAATATAAAGAAATTTCTAACATGAA 293  
Db 2988 AAATAAAAAAATCACTTAAACCTAAA 2963

RESULT 3  
US-10-172-086-18/c  
; Sequence 18, Application US/10172086  
; Publication No. US20030113750A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Method and nucleic acids for the differentiation  
; TITLE OF INVENTION: of prostate tumors  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/172,086  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 116  
; SEQ ID NO 18  
; LENGTH: 18997  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-172-086-18

Query Match 9.6%; Score 45.4; DB 15; Length 18997;  
Best Local Similarity 49.0%; Pred. No. 0.31;  
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
QY 99 CCAATAATTTATTTCCATGTCATGAGAGGACATGACATCCACAGAAAAATTTCTAATTAGCTTTGGT 218  
Db 11885 CAAAACCTATTTTAAACCCGTTATAACTAAATTTAAATTTACAACTAACTACCTTTC 11826  
QY 159 TAAACTCAATCAACAGGAGATGACATCCACAGAAAAATTTCTAATTAGCTTTGGT 218  
Db 11825 TCTCTCCATAAATACTAAATACTTATCATTAATAAATACTTAATAAATACTAT 11766  
QY 219 GTAGAAATTTGAACTGATACCTACATTAATTAATTAATTAATTAATTAATTAATTAATTA 278  
Db 11765 TAATAATAAACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 11706  
QY 279 AAGTCTTAACATGAAGCTAGTTCTTAACATGAAGACTAGTCCACGAACTCGTACCTTAT 338  
Db 11705 AACTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 11646  
QY 339 CCACAAA 345  
Db 11645 ACAAAA 11639

RESULT 4  
US-10-239-676-210/c  
; Sequence 210, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; 2000-04-06  
; 2000-04-07  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 210  
; LENGTH: 11812  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-210

Query Match 9.1%; Score 43.2; DB 15; Length 11812;  
Best Local Similarity 46.8%; Pred. No. 0.92; Mismatches 193; Indels 9; Gaps 1;  
Matches 178; Conservative 0;

QY 90 TTGACAGTCCAAATAATTTCCATGTCATGAGAGAGACATGATCAAGTAATAGC 149  
DB 1955 TTTTATTAAATTAATTTCTATATAACAAAATAATATAAAATTTTAA 1896

QY 150 TTAATCCCTAAATCAATCAACAGAGATGACACATCCACAGAAAATTTCTAATTAG 209  
DB 1895 TAATTTTATCAAAATATACGTTCAITCATACAAATACACAAATATATATACAT 1836

QY 210 TCTTGGCTGTAGAAATGGAAGTGAATACCTACATTAATCAAACTTTTGCAAAATAA 269  
DB 1835 CAATAATACATATTTTTCAAAATAAATAACGAATACATTTATATAACCACTCAAA 1776

QY 270 ATATAAAGAAAGTCTACATGAGAGCTAGTCTTAACATGAGACTAGTCCACGAATCG 329  
DB 1775 TTAATAATAAAGCATTAACAACATCCCTTTTTT-----TTTTTTTTTTTCT 1725

QY 330 TACCTTATCCACAAGGCTTAGACTTTCCAAATCGAGATTATCCCATGGACTGATGG 389  
DB 1724 TTTTTTTAAACAAAATCTTACTCTATAACCCAACTAAAATACATAACGCAATCTCG 1665

QY 390 ACACATCCAAATATCCCTATATAATACCTGCCATTCCTCCCTCCAGACTCATCTAAC 449  
DB 1664 ACTCACTAAATCCCACTCCTCTAAATTCAGTCACTTCTCCTCACTCAACCTCCGAATA 1605

QY 450 TCAAAAACAACACAAOCCA 469  
DB 1604 ACTAAACTACAAAGGCCCA 1585

## RESULT 5

US-09-754-853A-4/c  
; Sequence 4, Application US/09754853A  
; Publication No. US20030005491A1  
; GENERAL INFORMATION:  
; APPLICANT: Hauge, Brian M.  
; APPLICANT: Parnell, Laurence D.  
; APPLICANT: Parsons, Jeremy D.  
; APPLICANT: Wang, Ming Li  
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With  
; FILE REFERENCE: 38-10(15810)B  
; CURRENT APPLICATION NUMBER: US/09/754,853A  
; CURRENT FILING DATE: 2001-01-05  
; PRIOR APPLICATION NUMBER: US 60/174,880  
; PRIOR FILING DATE: 2000-01-07  
; NUMBER OF SEQ ID NOS: 1119  
; SEQ ID NO 4  
; LENGTH: 513509  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: {111805}..{113968}, {114684}..{115204}  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: {1}..{513509}  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: 318013\_region\_A3  
US-09-754-853A-4

Query Match 9.0%; Score 42.8; DB 12; Length 513509;  
Best Local Similarity 49.5%; Pred. No. 7.1;  
Matches 110; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 120 TGAGAGACGACATGACTAAAGTAATTAGCTTAATCCCTAAATCAATCAACAGAGA 179

DB 222302 TCAGAGGGAATCTTACAAAATTTGTCCTCTACTAGCCTACTAGTAACACACCTCCCTTA 22243

QY 180 TGACACATCCACAGAAAAAATCTAATTAAGTCTTTTCGCTGTAGAAAATTCGAAACTGAATA 239

DB 222242 CTACACTGATCCTAAACACCTTCTAATCCTGTGCTTAGTGGAATTTGGAACACACATT 222183

QY 240 CTTACATTAATTAACAACCTTTTGCAGATAAATAATAAAGAAAGTTCTTAACATGAAGACTAG 299

DB 222182 GCAAAATTAATAATAACCTTATGAATCAATAGAAATCAGAAACTTTCTATTATTAAATGC 222123

QY 300 TTTCAACATGAAGACATAGTCCAGAACTCGTACCTTATTCCCA 341

DB 222122 AGAAACATTAATCTCTGTACATAAATAATATATCAATTCCTCA 222081

## RESULT 6

US-10-239-676-96/c  
; Sequence 96, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003  
; CURRENT APPLICATION NUMBER: US/10/239,676  
; CURRENT FILING DATE: 2002-09-24  
; PRIOR APPLICATION NUMBER: PCT/EP01/03968  
; DE 10019058.8  
; DE 10019173.8  
; DE 10032529.7  
; DE 10043826.1  
; PRIOR FILING DATE: 2001-04-06  
; 2000-04-07  
; 2000-04-06  
; 2000-06-30  
; 2000-09-01  
; NUMBER OF SEQ ID NOS: 228  
; SEQ ID NO 96  
; LENGTH: 15732  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-96

Query Match 8.9%; Score 42.4; DB 15; Length 15732;  
Best Local Similarity 56.2%; Pred. No. 1.7;  
Matches 99; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 139 AAGTAATTAGCTTAATCCCTTAAACTCAATCAAAACGAGATGACACATCCACAGAAAA 198

DB 6621 AAATATTATACATAAAATCTTTAAACACTTCCACATATAAATCATATCACCAATAAAAA 6562

QY 199 ATTCTAATTAGTCTTTTCGCTGTAGAAAATTTGGAACCTGAATACCTAC-ATTAATTACAACT 257

DB 6561 AAAAAATTTACCTTTTCTTTCTTTTAAATATTTTATTTTCTTTATTTTACCTAT 6502

QY 258 TTGCAATAAATAATAAAGAAAGTTCTTAACATGAGAGACTAGTTCTTAACATGAAGA 313

DB 6501 TTTTAATAAAAAAATAAATAATTAATTAACCCCTCAATCTATCTCAACAAAAATA 6446

## RESULT 7

US-10-239-676-200/c  
; Sequence 200, Application US/10239676  
; Publication No. US20030082609A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
; FILE REFERENCE: 5013.1003

;; CURRENT APPLICATION NUMBER: US/10/239,676

;; CURRENT FILING DATE: 2002-09-24

;; PRIOR APPLICATION NUMBER: PCT/EP01/03968

;; DE 10019058.8

;; DE 10019173.8

;; DE 10032529.7

;; DE 10043826.1

;; PRIOR FILING DATE: 2001-04-06

;; 2000-04-06

;; 2000-04-07

;; 2000-06-30

;; 2000-09-01

;; NUMBER OF SEQ ID NOS: 228

;; SEQ ID NO 200

;; LENGTH: 4604

;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-200

Query Match 8.8%; Score 41.8; DB 15; Length 4604;

Best Local Similarity 47.2%; Pred. No. 1.4; Indels 0; Gaps 0;

Matches 127; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 206 TTACTCTTGGTGTAGAAATGTGAACACTGAATACCTACATTAATTACAACTTTTGCAAA 265

DB 3107 TTAATATATAAATTTAAATAATCAACACTACAAATCTAATACCCATTTTCAATAATCCCTA 3048

QY 266 TAAATATATAAAGAGTCTTAACATGAAGACTAGTCTTAACATGAAGACTAGTCCAGAA 325

DB 3047 TAAACTTTAAACAATAATTACATATAAATAAACAATTAAACTTAATCTTAACTTAATAA 2988

QY 326 CTCGTACCTATTCCCAAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTG 385

DB 2987 ATAATTCATTAATTCACAAACACTAAACACTACCTTAAACACCAACACTACATATACCA 2928

QY 386 ATGACACCCATCCAAATTTATCCCTTAAATACCTGCCCTCCCTCCCTCCACGACTATC 445

DB 2927 CATTAACGATTTAAATAATAATACATACACCAATCTTAACCTTTTAAATAATTT 2868

QY 446 TAACCTAAAAACAACACACACCAATCAT 474

DB 2857 AAACCTAAAAAAAACACGCACATATCGT 2839

#### RESULT 8

US-10-172-086-17/c

;; Sequence 17, Application US/10172086

;; Publication No. US20030113750A1

;; GENERAL INFORMATION:

;; APPLICANT: Epigenomics AG

;; TITLE OF INVENTION: Method and nucleic acids for the differentiation

;; TITLE OF INVENTION: of prostate tumors

;; FILE REFERENCE:

;; CURRENT APPLICATION NUMBER: US/10/172,086

;; CURRENT FILING DATE: 2002-06-13

;; NUMBER OF SEQ ID NOS: 116

;; SEQ ID NO 17

;; LENGTH: 18997

;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-172-086-17

Query Match

Best Local Similarity 8.6%; Score 41; DB 15; Length 18997;

Matches 107; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 99 CCAATAATTTATTCATGTCATGAGAGACATGACTAAAGTAATAGCTTAATCCCC 158

DB 10826 CCAATACCTACATTAATTAATCTAAACGTTACACACATAACAAAATAATTTCAATAAATAA 10767

QY 159 TAAAACTCAATACAAACGAGATGACACATCCACAGAAAAAATCTAATTAGTCTTTGGCGT 218

DB 10766 TTCAATAAAATACGAAAAACAACCAATACCAAAAAATRAATTAATAATAATAATAATA 10707

QY 219 GTAGAAATTTGGAACCTGAATACCTACATTAATTACAACTTTTGCAGAAATATAAAGA 278

DB 10706 AAAAAAATAAACAATTAATTATCTTTCAAAAAATCTATAATAAAAAAATAAATA 10647

QY 279 AAGTTCTAATGAAGACTAGTCTTAACATGAAGACT 315

DB 10646 AAATAATACTAAAAAATTTATCTATAANTCAACGAAT 10610

#### RESULT 9

US-10-239-676-24/c

;; Sequence 24, Application US/10239676

;; Publication No. US20030082609A1

;; GENERAL INFORMATION:

;; APPLICANT: OLEK, Alexander

;; APPLICANT: PIEPENBROCK, Christian

;; APPLICANT: BERLIN, Kurt

;; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

;; FILE REFERENCE: 5013.1003

;; CURRENT APPLICATION NUMBER: US/10/239,676

;; CURRENT FILING DATE: 2002-09-24

;; PRIOR APPLICATION NUMBER: PCT/EP01/03968

;; DE 10019058.8

;; DE 10019173.8

;; DE 10032529.7

;; DE 10043826.1

;; PRIOR FILING DATE: 2001-04-06

;; 2000-04-06

;; 2000-04-07

;; 2000-06-30

;; 2000-09-01

;; NUMBER OF SEQ ID NOS: 228

;; SEQ ID NO 24

;; LENGTH: 6158

;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-239-676-24

Query Match 8.8%; Score 40.8; DB 15; Length 6158;

Best Local Similarity 48.9%; Pred. No. 2.8;

Matches 137; Conservative 0; Mismatches 142; Indels 1; Gaps 1;

QY 190 ACAGAAAAAATCTAATTAGTCTTTGGGTGTAGAAATTTGGAACCTGAATACCTACATTAA 249

DB 5481 ATAAAAAAACTAAAACTAACCAATACATTTAACCAATTAACCAATTAACCAATTAACCA 5422

QY 250 TTACAACTTTTCG-AAATAAATATAAAGAAAGTTCTACACATGAAGACTAGTTCTTAACAT 308

DB 5421 AAAAAACTATTCCTAAATATTTTAAATTAATTTTAAATTAATTTTAAATTAATTTTAAAT 5362

QY 309 GAAGACTAGTCCAGCAACTCGTACCTTATTCACAAAGGCTTAGACTTTCCACAAATCGA 368

DB 5361 TATAATTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 5302

QY 369 GATTATCCCATGACATGATGGACACCATCCCAATTAATTCCTCTATAATACCTGCCCATCC 428

DB 5301 AAATACAAATAGCAAAATCTCAACTCACTACAACTCCACCTCCGAAATTCAGGCCATCT 5242

QY 429 CCTCTCCAGACTCATCTAACTCAAAACAACACACCAACC 468

DB 5241 CCTACCTCACTCCCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 5202

#### RESULT 10

US-10-239-676-30/c

;; Sequence 30, Application US/10239676



Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIORITY APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10033826.1  
PRIORITY FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 30  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-30

Query Match 8.5%; Score 40.2; DB 15; Length 7195;  
Best Local Similarity 46.6%; Pred. No. 4.3;  
Matches 129; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 140 AGTAATTAGCTTAATCCCTAAACTCAATAACAGAGATGACACATCCACAGAAAAA 199  
DB 1159 AATATAACACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1100  
QY 200 TTCTAATTAGCTTTGGGTGAGAAATGGAAATGAATACCTACATTAATACACTTT 259  
DB 1099 AACTACATCTCTTAATTTAATAAATAAATAAATAAATAAATAAATAAATAA 1040  
QY 260 TGCAATAAATAATAAGAAAGTTCTACATGAAGACTAGTCTTAACATGAAGACTGTC 319  
DB 1039 CCAATATATTTTAAACACATTTCTTAATCTTAAATAAATAAATAAATAAAT 980  
QY 320 CACGAATCTGACCTTATTCACAAAGCTTAGACTTTCCACAAATCGAGATTATCCAT 379  
DB 979 TAAATAAACTCTTAATAAACAATAAATAAATAAATAAATAAATAAATAAATAA 920  
QY 380 GGACTGATGGACACCATCCCAATATATCCCTATATA 416  
DB 919 TAAATCAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 883

RESULT 11  
US-09-352-6528/c  
Sequence 6528, Application US/09560352  
Patent No. US20020137139A1  
GENERAL INFORMATION:  
APPLICANT: Warren, Wesley C.  
APPLICANT: Tao, Nengbing  
APPLICANT: Byatt, John C.  
APPLICANT: Mathialagan, Nagappan  
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 6528  
LENGTH: 414  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 28-LIB3058-032-Q1-K1-G11

US-09-960-352-6528  
Query Match 8.4%; Score 39.8; DB 11; Length 414;  
Best Local Similarity 48.5%; Pred. No. 1.4;  
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 101 AATAATATTTCATGTCATGAGAGAGACGACATCCACAGAAAAATTTCTAGCTTTGCGTGT 160  
DB 350 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 291  
QY 161 AAATCAATACAAACGAGATGACATCCACAGAAAAATTTCTAGCTTTGCGTGT 220  
DB 290 AAATAAAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 231  
QY 221 AGAATTTGGAATGAATACCTACATTAATTAACACTTTGCAATAAATAAATAAATAA 280  
DB 230 AAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 171  
QY 281 GTTCTAATCAATGAAGACTAGTCTTAACATGAAGACTAGTCCAGGAAT 327  
DB 170 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 124

RESULT 12  
US-10-172-086-50/c  
Sequence 50, Application US/10172086  
Publication No. US20030113750A1  
GENERAL INFORMATION:  
APPLICANT: Epigenomics AG  
TITLE OF INVENTION: Method and nucleic acids for the differentiation  
FILE REFERENCE: of prostate tumors  
CURRENT APPLICATION NUMBER: US/10/172,086  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 50  
LENGTH: 6050  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-172-086-50

Query Match 8.4%; Score 39.8; DB 15; Length 6050;  
Best Local Similarity 46.5%; Pred. No. 5.1;  
Matches 128; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 100 CAATAATTTATTTCCATGTCATGAGAGAGACATGACTAAAGTAATAGCTTAATCCCT 159  
DB 4821 CACACACCTTTCACCTTAATATTAACATAACACCAACTAAACAATATACACTAAC 4762  
QY 160 AAACTCAATACAAACGAGATGACATCCACAGAAAAATTTCTAGCTTTGCGTG 219  
DB 4761 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4702  
QY 220 TAGAAATTTGAAACTGAATACCTTACATTAATTAACAATTTGCAATAAATAAATAA 279  
DB 4701 TCGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4642  
QY 280 AGTTCTAATCAATGAAGACTAGTCTTAACATGAAGACTAGTCCAGGAATCGTACCT 339  
DB 4641 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4582  
QY 340 CACAAGGCTTACACTTTCCACAAATCGAGATTAT 374  
DB 4581 TAAATATCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4547

RESULT 13  
US-09-790-988-1  
Sequence 1, Application US/09790988  
Patent No. US20020127687A1  
GENERAL INFORMATION:

APPLICANT: SHIGENOBU, SHUJI  
APPLICANT: WATANABE, HIDEMI  
APPLICANT: HATORI, MASAHIRA  
APPLICANT: SAKAKI, YOSHIYUKI  
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09/790,988  
CURRENT FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: JP2000-107160  
PRIOR FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Query Match 8.3%; Score 39.4; DB 11; Length 640681;  
Best Local Similarity 52.8%; Pred. No. 59;  
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 190 ACAGAAAATCTAAATTAGCTTTGCGTGTAGAAATGGAACTGGAATACCTACATTAA 249  
DB 611989 AATAAATATTGGAAGAAGTATGAGAAAATATTTTAAAAAATAATATAAGTAAAAA 612048

QY 250 TTACAACCTTTTGCAATATAATAAAGAAAGTTCTAACATGAAGACTAGTCTTAACATG 309  
DB 612049 TCATATGCTATTTTAAATAAATAGATAATTACATCTTATATAAAGTATTTTGTATTAGC 612108

QY 310 AAGACTAGTCCAGCACTCGTACCTTATTCACAAAGGCTT 350  
DB 612109 AAGAATATAAAGAAACAAAATATTATTTTAAGAGGCTT 612149

RESULT 14  
US-10-239-676-41/c  
Sequence 41; Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 41  
LENGTH: 6167  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-41

Query Match 8.2%; Score 39; DB 15; Length 6167;  
Best Local Similarity 49.7%; Pred. No. 8.3; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 276 AGAAGTTCTAACATGAGACTAGTCTTAACATGAAGACTAGTCCAGAACTCGTACCTT 335  
DB 2123 ATAAATAAACTAATATTAATATACCACATAAACCAGAACACACATACTCATACCTA 2064

QY 336 ATTCCACAAAGCGTTAGACTTTCCACAAATCGAGATTATCCATGGACTGATGGACACCA 395  
DB 2063 TATCCCAACACTTTAAAAAACCGGAACAAACGAAATCATATAAATCAAAAAATCGAAACCA 2004

QY 396 TCCAAATTCCTATATAATACCTGCCATTCCTCCCTCTCCAGACTCATCTAACTCAAAA 455  
DB 2003 TCCTAACTAACAGATAAAACCCCATCTCTACTAAAAATACAAAAAATTAACCGAAGT 1944

QY 456 ACAACACACACCAATCAT 474  
DB 1943 AATAATAACCACTATAAT 1925

RESULT 15  
US-09-938-842A-4290  
Sequence 4290; Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
TITLE OF INVENTION: SAME, AND METHODS OF USE  
FILE REFERENCE: SCRIPT300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 4290  
LENGTH: 864  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4290

Query Match 8.2%; Score 38.8; DB 11; Length 864;  
Best Local Similarity 64.4%; Pred. No. 3.6;  
Matches 58; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 193 GAAAAATTCCTAATAGTCTTTGCGTGTAGAAATGGAACTGAATACCTACATTAATTA 252  
DB 84 GAAAAAGATCTAAATAGGCTTTATCTGTAGACTTGGAGGTCAAAAACTAAAGTCTTTT 143

QY 253 CAACCTTTGCAATATAAATATAAAGAAAGT 282  
DB 144 AGACTCTTAGTAAATAAATGTTGAATAAGT 173

Search completed: July 20, 2003, 06:21:49  
Job time : 174 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2003, 04:50:04 ; Search time 1127 Seconds  
(without alignments)  
6825.959 Million cell updates/sec

Title: US-09-719-002-1  
Perfect score: 475  
Sequence: 1 gaattttattgcgaactga.....acaacacacacaaatcatg 475

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	61.2	12.9	1101	17	CNS0100X
c 2	59	12.4	1101	17	CNS0039G
c 3	48.4	10.2	1101	17	CNS0180R
c 4	47.4	10.0	1101	17	CNS0039G
c 5	45.6	9.6	1043	17	CNS0145P
c 6	45.2	9.5	1101	17	CNS0176E

7	44.2	9.3	1084	17	CNS0025S
8	44	9.3	358	9	AU262156
9	44	9.3	638	9	AL513901
c 10	43.8	9.2	860	17	CNS011EU
c 11	43.8	9.2	1101	17	CNS017P8
c 12	43.6	9.2	939	17	CNS00CNG
c 13	43.6	9.2	947	17	BH133700
c 14	43.4	9.1	767	17	CNS00AQX
c 15	43.4	9.1	919	17	CNS005RL
c 16	43.2	9.1	1101	17	CNS00L8X
c 17	43	9.1	301	17	CNS03C4J
c 18	43	9.1	848	17	CNS00GMQ
c 19	43	9.1	1086	17	CNS00YXK
c 20	43	9.1	1101	17	CNS000B8
c 21	43	9.1	1101	17	CNS0183Y
c 22	42.8	9.0	1201	17	CNS0167M
c 23	42.6	9.0	927	17	CNS00EU3
c 24	42.6	9.0	887	17	A2528056
c 25	42.6	9.0	1101	17	CNS00EVL
c 26	42.4	8.9	870	17	CNS06UZZ
c 27	42.4	8.9	1101	17	CNS007LP
c 28	42.2	8.9	924	17	CNS07A5L
c 29	42.2	8.9	1037	14	BQ648574
c 30	42.2	8.9	1039	17	CNS006H7
c 31	42.2	8.9	1077	17	CNS015CB
c 32	42.2	8.9	1101	17	CNS002FG
c 33	42	8.8	538	17	AQ523324
c 34	41.8	8.8	878	17	CNS0187R
c 35	41.8	8.8	1033	17	CNS0143L
c 36	41.4	8.7	300	17	BH872616
c 37	41.4	8.7	874	17	CNS007YZ
c 38	41.4	8.7	926	17	A2202638
c 39	41.4	8.7	1101	17	CNS00AY1
c 40	41.4	8.7	1101	17	CNS00LT2
c 41	41.2	8.7	512	14	C92998
c 42	41.2	8.7	622	17	DR7L8T
c 43	41.2	8.7	1038	17	CNS01TL7
c 44	41.2	8.7	1101	17	CNS0026Z
c 45	41	8.6	360	9	AU269640

#### ALIGNMENTS

RESULT 1	CNS0100X	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS0100X/c				
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN03G04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL098379	1	GI:5609990		
VERSION	AL098379.1				
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster.				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CGPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.				
FEATURES	Location/Qualifiers				



```

source      1. .l101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_="BACN37D08"
            /clone_lib="DrosBAC"
            /plasmid="pBelobAC11"
            /notes="end : SP6"

BASE COUNT   329 a    205 c   116 g    304 t   147 others
ORIGIN

Query Match          10.2% Score 48.4; DB 17; Length 1101;
Best Local Similarity 34.2%; Pred. No. 0.21;
Matches 104; Conservative 55; Mismatches 145; Indels 0; Gaps 0;

QY      165 TCAATACAGCAGAGTGACACATCCACAGAAATAATTCTAATTAGTCTTTGGCTGTAGAA 224
         ||||| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db       717 TTAATWAATWATGNTATTTATATATAAAAAAMWTATGTCATCTCTTIYYANTNNWTT 776

QY      225 ATTGGAACATGAATACCTACATTAATTACAACATTTTGCAAAATAAAATATAAGAAAGTTC 284
         ||||| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db       777 ATAAAWAAAMWHCMAAAATAATTTTTTCTAAATCTCTAATAWAHYHTWAAAAAWMHATT 836

QY      285 TAACATGAAGACTAGTICTAACATGATGAAGACTAGTCCACGAACCTCGTACCCTTATTCACAA 344
         ||||| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db       837 TIYTCTTWATMTTAAAWAHATWMTTWAAAWAAAAAAAAMAATWWAAAAHMTCAAAM 896

QY      345 AGGCTTAGACTTTCCACAAATCAGAGATTATCCCATGGACTGATGGACACCATCCAATTA 404
         ||||| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db       897 AAAATYMTCTCTMTCTCTATYCAYCTYTTCTCTYATGCCMTAWAYCCTMTTAAMAAHA 956

QY      405 TCCTATAATACCTGCCCATTCCTCCCTCCCTCCAGACTCATCTAACTCAAAACCAACACAC 464
         ||||| : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db       957 ATAAAAAAAANAAMMYMWCWCTMTTYAHCTCTWMYCYTCTCTCTCMWCAAAAAAAMA 1016

QY      465 AACG 468
         :::|
Db       1017 MMNC 1020

```

RESULT 4  
CNS0039G/c  
LOCUS  
DEFINITION  
BACR08K10 of RPCI-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.  
ACCESSION  
AL063921  
VERSION  
AL063921.1  
KEYWORDS  
GI:4941778  
SOURCE  
GSS.  
ORGANISM  
*Drosophila melanogaster*.  
*Drosophila melanogaster*.  
*Eukaryota*; *Metazoa*; *Arthropoda*; *Hexapoda*; *Insecta*; *Pterygota*; *Neoptera*; *Endopterygota*; *Diptera*; *Brachycera*; *Muscomorpha*; *Ephyroidea*; *Drosophilidae*; *Drosophila*.  
REFERENCE  
1 (bases 1 to 1101)  
AUTHORS  
Genoscope.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
COMMENT  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
- Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley *Drosophila* Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see [http://www.fruitfly.org/TheBDGP/Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pierer de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

```

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1. .1101
   /organism="Drosophila melanogaster"
   /db_xref="taxon:72227"
   /clone="BACR08K10"
   /clone_lib="RPCi-98"
   /note="end : TET3"
BASE COUNT      201 a    64 c    131 g    202 t    503 others
ORIGIN
Query Match          10.0% ; Score 47.4; DB 17; Length 1101;
Best Local Similarity 26.7%; Pred. No. 0.36; Indels 1; Gaps 1;
Matches 92; Conservative 104; Mismatches 148;
Qy 92 GACAGTTCCCAATTAATTTCATGTGCATGAGAGACACATGACTAAAGTAACTTAGCTT 151
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 682 RDRATWDRDTDAWADAAWTTTDDTDDBDKRRRRRGARRRRTJARAADWDWTWKAMD 623
Qy 152 AATCCCCATAAATCAATACAAGGAGATGACACATCCACAGAAAAATTCATATAGTC 211
   :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 622 WAKDWKTTRADRDRWAARDTWTD-ARKADRDWAKARAWRAARDRAARADRWTTKGKT 564
Qy 212 TTTCGGTGTAAGAATCGGAACAGTACCTACATTAATACAACCTTTTCGAATAAAAT 271
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 563 TTATITWAARAAWWANWATTTATTTTTTTTWTWTWTWTWTWTWTWTAAWAAWWT 504
Qy 272 ATAAAGAAAGTCTTCAATGAAGACTAGTTCTTAACATGAAGACTAGTCCACGAACTCGTA 331
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 503 ATWAAATAAWAAAAAAAATAATTTTTTTTTTTTWTAAWTAATAWTTTWTWTWTWTWAA 444
Qy 332 CCTATTCCCAAGGCTTAGACTTCCACAAATCGAGATTATCCCATGGCATGATGGAC 391
   |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 443 TTTTITWTWTWAAATTTATTTTWTWTAAAAAAAWWWWTATATKCCCCCVCVCCC 384
Qy 392 ACCATCCAAATATFCCCTATAAATACCTGCCATTCGCCCTCCCTC 436
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 383 CCYCCGCCCTGCYCTCTCTCTGTCGCVVVVVVVVVVVVVVVVVVVVVVQCVC 339

```

RESULT 5	
CNS0145P/c	
LOCUS	1043 bp DNA linear GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN1g11 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL103735
VERSION	AL103735.1 GI:5615346
KEYWORDS	GSS.
ORGANISM	Drosophila melanogaster. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1043)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (23-JUL-1999), Genoscope - Centre National de Sequencage :
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES	Location/Qualifiers
source	1..1043 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN1g11"

```

/clone_lib="DrosBAC"
/clone_lib="pBelobAC11"
/notes="end : T7"
BASE COUNT 277 a 96 c 121 g 382 t 167 others
ORIGIN
Query Match 9.6%; Score 45.6; DB 17; Length 1043;
Best Local Similarity 36.4%; Pred. No. 0.96;
Matches 94; Conservative 41; Mismatches 123; Indels 0; Gaps 0;
QY 51 AAATTTCTGTGGCCACACACTGCTCCTGCTGATTCACAGTTCCTCAATATTT 110
Db 840 ATATWTARGGRAADRAAAAAAATTTTWTWMMWMTTWWMMWMTTWTWT 781
QY 111 TCCATGCTGAGAGAGACATGACTAAAGTAATAGCTTAATCCCTTAAACTCAATA 170
Db 780 TTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 721
QY 171 CAACGAGATGACATCCACAGAAAAATTTCTAATAGTCTTTCGGCTGAGAAATGGA 230
Db 720 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 661
QY 231 AACTGAATCACTACATTAATTAACACTTTTGGCAATAAATAAATAAAGAAAGTTCTA 290
Db 660 WAAAAATATATTTTWWATATAAATAAATAAATAAATAAATAAATAAATAAATAA 601
QY 291 GAAGACTAGTCTTAACAT 308
Db 600 TAAAAATATATATATATATATATATATATATATATATATATATATATATATATAT 583

RESULT 6
CNS0176E 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN17D02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL107648
VERSION AL107648.1 GI:5627952
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT SubMITTED (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES
source Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACN17D02"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : T7"
BASE COUNT 337 a 143 c 131 g 350 t 140 others
ORIGIN
Query Match 9.5%; Score 45.2; DB 17; Length 1101;
Best Local Similarity 32.7%; Pred. No. 1.2;
Matches 114; Conservative 63; Mismatches 172; Indels 0; Gaps 0;
QY 51 AAATTTCTGTGGCCACACACTGCTCCTGCTGATTCACAGTTCCTCAATATTT 110
Db 840 ATATWTARGGRAADRAAAAAAATTTTWTWMMWMTTWWMMWMTTWTWT 781
QY 111 TCCATGCTGAGAGAGACATGACTAAAGTAATAGCTTAATCCCTTAAACTCAATA 170
Db 780 TTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 721
QY 171 CAACGAGATGACATCCACAGAAAAATTTCTAATAGTCTTTCGGCTGAGAAATGGA 230
Db 720 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 661
QY 231 AACTGAATCACTACATTAATTAACACTTTTGGCAATAAATAAATAAAGAAAGTTCTA 290
Db 660 WAAAAATATATTTTWWATATAAATAAATAAATAAATAAATAAATAAATAAATAA 601
QY 291 GAAGACTAGTCTTAACAT 308
Db 600 TAAAAATATATATATATATATATATATATATATATATATATATATATATATATAT 583

RESULT 7
CNS0025S 1084 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN01E02 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097258
VERSION AL097258.1 GI:5608869
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT SubMITTED (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES
source Location/Qualifiers
1..1084
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACN01E02"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/notes="end : T7"
BASE COUNT 453 a 186 c 117 g 173 t 155 others
ORIGIN
Query Match 9.3%; Score 44.2; DB 17; Length 1084;
Best Local Similarity 35.2%; Pred. No. 2.1;
Matches 127; Conservative 59; Mismatches 173; Indels 2; Gaps 1;
QY 97 TTCCATTAATATTTTCCATGTCATGACAGAGACATGACTAAAGTAATAGCTTAATCC 156
Db 724 TTTTAAAAAAAAMCYCTTTTAAAAAAAAMCYCTTTTAAAAAAAAMCYCTTTTAAAAAAA 783

```

```

QY 157 CTTAACTCTACATCAACGAGATGACATCCACAGAAAAAATCTAATTAGTCTTGC 216
Db 784 TAAAAAATAATTTAAACCTTTAAAAAATAAATAAATAAATAAATAAATAAATAA 843
QY 217 GTGTAGAAATGGAACTGAATCTACATTAATTACAACCTTTGCAATAAATAAATAA 276
Db 844 AWTAAAAAATGCTGTTTAAAAAATAAATTTTAAATTTTAAATTTAAATAAATAA 903
QY 277 GAAAGTCTACATGACAGACTAGTCTTAACATGAGACTAGTCCACGAACCTGACTTA 336
Db 904 WWWWWWWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 961
QY 337 TTCCAAAGAGCTTAGACTTTCCCAAAATCGAGATTTATCCCATGGAAGTGGACACCAT 396
Db 962 TWMAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 1021
QY 397 CCAATTTATCCCTATTAATCTGCTCCATTCCTCTCTCCAGACTCATCTAATCAAAA 456
Db 1022 ATAHAAMMCMCTAWAAAAAAMMCCCAAAAMMAACAWMCTACMMAMCMCMCCMAMM 1081
QY 457 C 457
Db 1082 C 1082

RESULT 8
LOCUS AU262156 358 bp mRNA linear EST 09-MAY-2002
DEFINITION AU262156 VS Dictyostelium discoideum cDNA clone VSB644 3', mRNA
sequence.
ACCESSION AU262156.1 GI:20519703
KEYWORDS Dictyostelium discoideum.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 358)
AUTHORS Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
Urushihara,H., Morio,T., Saito,T., Koriki,E., Ochiai,H., Maeda,M.,
Takeuchi,I., Kohara,Y. and Tanaka,Y.
TITLE Population analysis of cDNAs from unicellular and multicellular
stages of Dictyostelium discoideum
JOURNAL Unpublished (2002)
COMMENT Contact: Hideo Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideo@biol.tsukuba.ac.jp.
FEATURES
source
1. 358
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="VSB644"
/clone_lib="VS"
/sex="mat A"
/dev_stage="vegetative"
BASE COUNT 167 a 37 c 56 g 98 t
ORIGIN
Query Match 9.3%; Score 44; DB 9; Length 358;
Best Local Similarity 52.1%; Pred.No.2.3;
Matches 98; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 92 GACAGTTCATTAATTTTCCATGTCATGAGAACGACATGACTAAAGTAATAGCTT 151
Db 158 GTCAATTTTGAACAUGCTTCATGTATGTATGTAATGTGTGGACCAAAATATTGCTC 227
QY 152 AATCCCTTAAACTCAATACAAAGAGATGACATCCACAGAAAAAATCTAATTAGTC 211
Db 228 AATTAAGCTAAATATATGCGGATATGCTGAAATAAGTATATAAATAAGTATATAA 287

```

```

QY 212 TTTCGCTGTAGAATTTGGAACTGAATACCTACATTAATTACAACTTTTGCATAAAT 271
Db 288 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 347
QY 272 ATAAAGAA 279
Db 348 ATTTAAAA 355

RESULT 9
LOCUS AL513901 638 bp mRNA linear EST 13-FEB-2001
DEFINITION AL513901 LIT_NFL006_PL2 Homo sapiens cDNA clone CLOBA0062G01 3
prime, mRNA sequence.
ACCESSION AL513901
VERSION AL513901.1 GI:12777395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 638)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. 638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA0062G01"
/clone_lib="LIT_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 347 a 58 c 15 g 141 t 77 others
ORIGIN
Query Match 9.3%; Score 44; DB 9; Length 638;
Best Local Similarity 39.5%; Pred.No.2.3;
Matches 98; Conservative 31; Mismatches 119; Indels 0; Gaps 0;

QY 119 ATGAGAGAACGACATGACTAAAGTAATAGCTTAAATCCCTTAAATCAATCAACAGAG 178
Db 174 ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 233
QY 179 ATGACATATCCACAGAAAAATCTAATAGCTTTTGGTGTAGAAATTTGGAACCTGAAT 238
Db 234 GGGAAAAATTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 293
QY 239 ACCTACATTAATTTACAACCTTTTGCATAAATAAATAAATAAATAAATAAATAA 298
Db 294 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 353
QY 299 GTTCTACATGAGACTAGTCCACGAGACTCGTACCTTATCCACAAGGCTTAGCTTC 358
Db 354 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 413
QY 359 CACAAATC 366
Db 414 AAAAAACC 421

```

```

RESULT 10
CNS011EU/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence SP6 end of BAC
  BACN06H18 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL100176.1 GI:5611787
VERSION
  1
KEYWORDS
  GSS.
ORGANISM
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 860)
REFERENCE
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
FEATURES
  Location/Qualifiers
  source
  1..860
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone_lib="DrosBAC"
  /plasmid="pBelobAC11"
  /note="end : SP6"
BASE COUNT
  202 a 106 c 116 g 306 t 130 others
ORIGIN
  Query Match
  Best Local Similarity 9.2%; Score 43.8; DB 17; Length 860;
  Matches 131; Conservative 27; Mismatches 157; Indels 3; Gaps 1;
  QY 54 TTTCGTGCGGCACACATACCTGTCCTTGTGTTGACAGTCCCAATATATTTC 113
  DB 319 TTNTGACNCGGWTTCNATRCWCGCGTTCNATATATATGCGCGT---AATPAAGANTGAA 263
  QY 114 ATGTCATGAGAGACATGACTAAAGTAATAGCTTAATCCCTCAAACTCAATACAA 173
  DB 262 ATGAAAAAATAGCAAMATRGAAAAAGAAATCTCTTTTCTCATWCMAGAAWCTTC 203
  QY 174 ACGAGATGACATCCACAGAAAAATTCATAGTCTTTGCGGTAGAAATTTGGAAC 233
  DB 202 WMMMAAAAAAATAAAGAAATAAAAAAATCTASATTTKACAATKCTTAWWWTAAAA 143
  QY 234 TGAATACCTTACATTAATTACACATTTTGCAATAAATAATATAAGAAAGTTCTACATCAA 293
  DB 142 AMAAKAATAAATAATTAATAAAWKNNAATTTATATAAATWAATGTTTAAATATAA 83
  QY 294 GACTAGTCTACATGAGACTAGTCCAGCACTGCTACCTTATCCCAAGGCTTAGA 353
  DB 82 AGATAGTATATAATATTCCTTTGTTTAKAAMTTTAHAWHTTAATATATTTAATTAAGT 23
  QY 354 CTTTCACAAAATCGAGAT 371
  DB 22 CATTTAATTATATGAT 5
RESULT 11
CNS017PB/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence SP6 end of BAC
  BACN37B03 of DrosBAC library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL108329.1 GI:5628633
VERSION
  1
KEYWORDS
  GSS.
ORGANISM
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 1101)
REFERENCE
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
FEATURES
  Location/Qualifiers
  source
  1..1101
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone_lib="DrosBAC"
  /plasmid="pBelobAC11"
  /note="end : SP6"
BASE COUNT
  252 a 145 c 148 g 371 t 185 others
ORIGIN
  Query Match
  Best Local Similarity 9.2%; Score 43.8; DB 17; Length 1101;
  Matches 117; Conservative 14; Mismatches 139; Indels 0; Gaps 0;
  QY 69 ACATATGGTCCTTGCTTGATTGACAGTTCCTCAATATATTTCCTGTCATGAGAGAAG 128
  DB 299 ACAGATTAGCAGTGGCACAAATAAATAAGATAATAATAATAATAATAATAATAA 240
  QY 129 CACATGACTAAAGTAATAGCTTAATCCCTAAAGTCAATACAAACGAGATGACACATC 188
  DB 239 MACAAAGAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAATAA 180
  QY 189 CACAGAAAAAATTTCTAATTAGTCTTTGCGGTAGAAATTTGGAACCTGACTCATTA 248
  DB 179 AAAAAAAMAAAAAAGAAAGWAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA 120
  QY 249 ATTACACATTTTCCAAATAAATAATAAAGAAAGTTCTAATCAATGAAGACTAGTTCTAACAT 308
  DB 119 AAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 60
  QY 309 GAAGACTAGTCCACGAACTGCTACCTTAT 338
  DB 59 CATCATCTNTTCAAGTTATCTGTTTTTATT 30
RESULT 12
CNS00CNG
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL059400.1 GI:4945964
VERSION
  1
KEYWORDS
  GSS.
ORGANISM
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  1 (bases 1 to 939)
REFERENCE
  Direct Submission
  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the European Drosophila Genome Project (EDGP) -
  http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
  library (Dros BAC) was made by Alain Billaud at CEPH (Centre
  d'Etude du Polymorphisme Humain) with funding provided by a MRC
  project grant. The DNA was prepared from embryos by Alain Bucheton
  and Genevieve Payan. It has been constructed in the vector
  pBelobAC11.
FEATURES
  Location/Qualifiers
  source
  1..939
  /organism="Drosophila melanogaster"
  /db_xref="taxon:7227"
  /clone_lib="DrosBAC"
  /plasmid="pBelobAC11"
  /note="end : SP6"
BASE COUNT
  252 a 145 c 148 g 371 t 185 others
ORIGIN
  Query Match
  Best Local Similarity 9.2%; Score 43.8; DB 17; Length 1101;
  Matches 117; Conservative 14; Mismatches 139; Indels 0; Gaps 0;
  QY 69 ACATATGGTCCTTGCTTGATTGACAGTTCCTCAATATATTTCCTGTCATGAGAGAAG 128
  DB 299 ACAGATTAGCAGTGGCACAAATAAATAAGATAATAATAATAATAATAATAA 240
  QY 129 CACATGACTAAAGTAATAGCTTAATCCCTAAAGTCAATACAAACGAGATGACACATC 188
  DB 239 MACAAAGAAAAAATAAATAATAATAATAATAATAATAATAATAATAATAA 180
  QY 189 CACAGAAAAAATTTCTAATTAGTCTTTGCGGTAGAAATTTGGAACCTGACTCATTA 248
  DB 179 AAAAAAAMAAAAAAGAAAGWAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAA 120
  QY 249 ATTACACATTTTCCAAATAAATAATAAAGAAAGTTCTAATCAATGAAGACTAGTTCTAACAT 308
  DB 119 AAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAA 60
  QY 309 GAAGACTAGTCCACGAACTGCTACCTTAT 338
  DB 59 CATCATCTNTTCAAGTTATCTGTTTTTATT 30

```





The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# FEATURES

source

1. .767

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR12A20"

/clone\_lib="RPCI-98"

/note="end : TET3"

BASE COUNT 312 a 105 c 65 g 164 t 121 others

Query Match 9.1%; Score 43.4; DB 17; Length 767;

Best Local Similarity 36.2%; Pred. No. 3.2; Mismatches 96; Indels 0; Gaps 0;

Matches 77; Conservative 40; Mismatches 96; Indels 0; Gaps 0;

QY 113 CATGTCATGAGAGACATGCTAAGTAATAGTCTTATCCCTTAAACCAATACATA 172

DB 135 MAATAAAMWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 194

QY 173 AACGAGATGACATCCACAGAGAAATCTCTAGTCTTGGTGTGAGAAATGGAAA 232

DB 195 ATAATATATATATTTCAATACACWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 254

QY 233 CTGATACCTACATTAATACATCTTTGCAATATAAATAAAGTAAGTCTTAACATGA 292

DB 255 AAAAAAAMWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 314

QY 293 AGACTAGTCTAACATGAAGACTAGTCCAGAA 325

DB 315 WTWMAWTTACWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAW 347

# RESULT 15

CNS005RL

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS005RL 919 bp DNA linear GSS 03-JUN-1999

Drosophila melanogaster genome survey sequence T7 end of BAC #

BACR12F23 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

AL061409

GSS.

AL061409.1 GI:4943512

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 919)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr))

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammos in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# FEATURES

source

1. .919

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="BACR12F23"

/clone\_lib="RPCI-98"

/note="end : T7"

BASE COUNT 279 a 124 c 93 g 197 t 226 others

ORIGIN

Query Match 9.1%; Score 43.4; DB 17; Length 919;

Best Local Similarity 22.8%; Pred. No. 3.2; Mismatches 163; Indels 1; Gaps 1;

Matches 85; Conservative 123; Mismatches 163; Indels 1; Gaps 1;

QY 99 CCAATAATATTATTCATGTGTCATGAGAGACGACATGACTAAAGTAATAGCTTAATCCCC 158

DB 521 CAAAAAATCAACACATWACCTTTTCAATTCCTTCCCAATWCTATWAAWWHHHHAHA 580

QY 159 TAAACTCAATCAACAGAGATGACATCCACAGAAAAAATCTAA-TTACTCTTTGGG 217

DB 581 MHHMMCMWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 640

QY 218 TGTAGAAATTTGGAACCTGAATACCTACATTAATTTACAACTTTTGCAGAAATAAAG 277

DB 641 TXHACATATAAAAAACAAATWTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 700

QY 278 AAGTCTCAACATGAAGACTAGTCTTAACATGAAGACTAGTCCACGAACTCGTACTTAT 337

DB 701 WMMNAHMMWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 760

QY 338 TCCACAAAGGCTTAGACTTTCCACAAATCGAGATTATCCCATGGACTGACACCATC 397

DB 761 WAWTTWAANWCMWMTATTTTMMWCMWCMWCMWCMWCMWCMWCMWCMWCMWCMW 820

QY 398 CAATATATCCCTATAATACCTGCGCATTCGCCCTCCCTCCAGACTCATCTTAACAAAC 457

DB 821 MAWCCMACMMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 880

QY 458 AACACACAACCA 469

DB 881 MWTWTHHHMW 892

Search completed: July 20, 2003, 06:08:28

Job time : 1131 secs